Université de Lille

Doctoral School **ED Régionale SMRE 104**

University Department **Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM)**

Thesis defended by **Georges Murr**

Defended on **December 20, 2024**

In order to become Doctor from Université de Lille

Academic Field **Physics**

Speciality **Dilute Media and Fundamental Optics**

Machine Learning-Assisted Spatiotemporal Chaos Forecasting

Thesis supervised by Saliya Coulibaly

Committee members

Referees	Bertrand KIBLER	Professor at Université de Bourgogne	
	Stefania RESIDORI	CNRS - CEO de la société HOASYS	
Examiners	Arnaud MUSSOT	Professor at Université de Lille	Committee President
	Marcel CLERC	Professor at University of Chile	
	Mustapha TLIDI	Professor at Université Libre de Bruxelles	
Supervisor	Saliya Coulibaly	HDR Associate Professor at Université de Lille	

Université de Lille

Université de Lille

Doctoral School **ED Régionale SMRE 104**

University Department **Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM)**

Thesis defended by **Georges Murr**

Defended on **December 20, 2024**

In order to become Doctor from Université de Lille

Academic Field **Physics**

Speciality **Dilute Media and Fundamental Optics**

Machine Learning-Assisted Spatiotemporal Chaos Forecasting

Thesis supervised by Saliya Coulibaly

Committee members

Referees	Bertrand KIBLER	Professor at Université de Bourgogne	
	Stefania RESIDORI	CNRS - CEO de la société HOASYS	
Examiners	Arnaud MUSSOT	Professor at Université de Lille	Committee President
	Marcel CLERC	Professor at University of Chile	
	Mustapha TLIDI	Professor at Université Libre de Bruxelles	
Supervisor	Saliya Coulibaly	HDR Associate Professor at Université de Lille	

UNIVERSITÉ DE LILLE

École doctorale **ED Régionale SMRE 104**

Unité de recherche **Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM)**

Thèse présentée par **Georges MURR**

Soutenue le **20 décembre 2024**

En vue de l'obtention du grade de docteur de l'Université de Lille

Discipline **Physiques**

Spécialité **Milieux Dilués et Optique Fondamentale**

Optimisation de l'Horizon de prédictibilité des Evènements Extrêmes par «Deep Learning»

Thèse dirigée par Saliya COULIBALY

Composition du jury

Rapporteurs	Bertrand KIBLER	professeur à l'Université de Bourgogne	
	Stefania RESIDORI	CNRS - CEO de la société HOASYS	
Examinateurs	Arnaud MUSSOT	professeur à l'Université de Lille	président du jury
	Marcel CLERC	professeur à l'University of Chile	
	Mustapha TLIDI	professeur à l'Université Libre de Bruxelles	
Directeur de thèse	Saliya COULIBALY	MCF HDR à l'Université de Lille	

Keywords: complex systems, spatiotemporal chaos, turbulence, machine learning, artificial neural networks, information theory

Mots clés : systèmes complex, chaos spatiotemporel, turbulence, apprentissage automatique, réseaux de neurones artificiels, théorie de l'information

This thesis has been prepared at

Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM)
CNRS UMR 8523
Université de Lille
Bâtiment P5 campus Cité Scientifique
2 Avenue Jean Perrin
59655 Villeneuve d'Ascq Cedex
France

☎	(33) 03 20 33 60 37
🖂	phlam-contact@univ-lille.fr
Web Site	https://phlam.univ-lille.fr/

Prediction is very difficult,
especially if it's about the future.

Niels Bohr

Life is divided into three terms - that which was, which is, and which will be. Let us learn from the past to profit by the present, and from the present, to live better in the future.

William Wordsworth

Machine Learning-Assisted Spatiotemporal Chaos Forecasting

Abstract

From towering rogue waves to powerful winds, extreme events can disrupt natural systems and human activity without warning. Though seemingly unpredictable, these events often arise from the complex dynamics of chaotic systems, particularly spatiotemporal chaos, where patterns unfold across both time and space. In this thesis, we study extreme events in optical systems, focusing on an optical fiber ring resonator modeled by the Lugiato-Lefever equation. This setup provides a controlled environment to analyze the chaotic behaviors that lead to such phenomena. Recent advancements in machine learning, especially neural networks, offer new tools for predicting chaotic dynamics. However, long-term forecasting remains challenging due to chaos's inherent unpredictability. We propose extending the prediction horizon using information theory methods, like transfer entropy, to identify local regions contributing to extreme events and improve forecast accuracy. Additionally, we examine the turbulent dynamics generated by solitons in these systems, providing explanations for their onset and evolution. Our analysis offers new insights into chaotic behavior. Finally, we propose applying these methods to real-world wind dynamics to enhance forecasting and deepen understanding of chaotic natural systems.

Keywords: complex systems, spatiotemporal chaos, turbulence, machine learning, artificial neural networks, information theory

---

Optimisation de l'Horizon de prédictibilité des Evènements Extrêmes par «Deep Learning»

Résumé

Des vagues scélérates aux vents violents, les événements extrêmes peuvent perturber les systèmes naturels et les activités humaines sans avertissement. Bien que ces événements semblent imprévisibles, ils émergent souvent des dynamiques complexes des systèmes chaotiques, en particulier du chaos spatiotemporel, où des motifs se déploient dans le temps et l'espace. Dans cette thèse, nous étudions les événements extrêmes dans des systèmes optiques, en nous concentrant sur un résonateur à fibre optique modélisé par l'équation de Lugiato-Lefever. Ce système offre un environnement contrôlé pour analyser les comportements chaotiques à l'origine de ces phénomènes. Les récents progrès en apprentissage automatique, notamment avec les réseaux de neurones, offrent de nouveaux outils pour prédire les dynamiques chaotiques. Cependant, la prévision à long terme reste difficile en raison de l'imprévisibilité inhérente au chaos. Nous proposons d'étendre l'horizon de prédiction en utilisant des méthodes de théorie de l'information, telles que l'entropie de transfert, pour identifier les régions locales contribuant aux événements extrêmes et améliorer la précision des prévisions. En outre, nous examinons les dynamiques turbulentes générées par les solitons dans ces systèmes, en proposant des explications sur leur apparition et leur évolution. Notre analyse offre de nouvelles perspectives sur le comportement chaotique. Enfin, nous proposons d'appliquer ces méthodes aux dynamiques du vent en situation réelle pour améliorer les prévisions et approfondir la compréhension des systèmes chaotiques naturels.

Mots clés : systèmes complex, chaos spatiotemporel, turbulence, apprentissage automatique, réseaux de neurones artificiels, théorie de l'information

---

Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM)
CNRS UMR 8523 – Université de Lille – Bâtiment P5 campus Cité Scientifique – 2 Avenue Jean Perrin – 59655 Villeneuve d'Ascq Cedex – France

Acknowledgement

Finishing this PhD has been one of the most challenging and rewarding experiences of my life. It has been a long journey full of learning, growth, discovery, and resilience. I am deeply grateful to everyone who has supported me along the way.

First, I would like to thank the members of my jury: Bertrand Kibler, Stefania Residori, Arnaud Mussot, Marcel Clerc, and Mustapha Tlidi. I truly appreciate the time and effort they dedicated to evaluating my work, as well as their insightful feedback, which has enriched this thesis.

A special thanks to Marcel Clerc, who welcomed me during my visit to Chile. His generosity, along with the warm hospitality of his team, made my stay both productive and memorable.

I owe an immense debt of gratitude to my supervisor, Saliya Coulibaly, for his guidance, patience, and countless valuable discussions. His support has shaped this work in more ways than I can count. And of course, the number of beers I owe him… well, let's just say it's a long tab.

I have been lucky to share this journey with incredible colleagues and friends. Their advice, encouragement, and occasional distractions made the process far more enjoyable.

To my family, thank you for your unwavering love, support, and prayers. Your belief in me has been my greatest strength.

Finally, my deepest gratitude goes to Fadia, my fiancée. Through every high and low, she has been there, offering endless support and encouragement. Without her, I wouldn't have made it to the finish line.

To all of you, thank you from the bottom of my heart.

Table of Contents

List of Tables

List of Figures

Acronyms

D_{KY}	Kaplan-Yorke Dimension
ζ	Median Symmetric Accuracy
AI	Artificial Intelligence
ANN	Artificial Neural Networks
CGLE	Complex Ginzburg-Landau Equation
CMI	Conditional Mutual Information
CS	Cavity Solitons
CW	Continuous-Wave
DL	Deep Learning
DNN	Deep Neural Network
DS	Dynamical System
EE	Extreme Events
FDM	Finite Difference Method
FNN	False Nearest Neighbors
FTP	Fast time periodicity
GRU	Gated Recurrent Unit
HSS	Homogeneous Steady State
JIDT	Java Information Dynamics Toolkit
KFC	Kerr Frequency Comb
KSE	Kuramoto-Sivashinsky Equation
KSG	Kraskov-Stögbauer-Grassberger
LE	Lyapunov Exponent
LLE	Lugiato-Lefever Equation
LS	Lyapunov Spectrum

LSA	Linear Stability Analysis
LSTM	Long Short-Term Memory
MAE	mean absolute error
MAX	Maximum
MI	Modulation Instability
ML	Machine Learning
MSE	mean squared error
NDTE	Normalized Directionality Transfer Entropy
NLLE	Nonlinear Local Lyapunov Exponent
NLSE	Nonlinear Schrödinger Equation
NN	Neural Network
ODE	Ordinary Differential Equations
OFC	Optical Frequency Combs
PC	Pearson correlation
ReLU	Rectified Linear Unit
RGIE	Relative Growth of the Initial Error
RK4	Fourth Order Runge-Kutta
RMSE	Root mean squared error
RNN	Recurrent Neural Networks
sMAPE	Symmetric Mean Absolute Percentage Error
STC	Spatiotemporal Chaos
STD	Standard deviation
STP	Slow time periodicity
TE	Transfer Entropy
WD	Wind Direction
WS	Wind Speed

Introduction

The unpredictability of nature has long captured human curiosity, driving the development of models to forecast the unexpected. From powerful hurricanes to unexpected financial collapses, extreme events often appear without warning, causing widespread damage to communities and the environment. But what if we could anticipate such events with more precision? What if we could peer into the chaotic underpinnings of these phenomena and find a way to predict them before they strike?

Extreme events are not just the products of random chance; they are deeply rooted in the laws of chaotic dynamics, where the delicate balance of complex systems can be disrupted by even the smallest of disturbances. These systems are governed by what is often referred to as the "butterfly effect"—small changes in initial conditions can lead to dramatically different outcomes [1]. This sensitivity to initial conditions makes forecasting these systems challenging, as small uncertainties in measurement can quickly escalate into wildly different predictions. Despite their apparent randomness, chaotic systems follow deterministic rules, creating complex patterns that give rise to the extreme events we experience in the real world. Consider the real-world impact of these extreme events: hurricanes that flatten entire cities, wildfires that decimate forests, financial markets that crash without warning, and rogue waves that unexpectedly arise in the ocean, threatening ships and offshore structures. Their ability to disrupt economies and upend lives has made them a central focus of research across many disciplines. In fields ranging from meteorology to finance, scientists strive to understand how these events emerge, with the ultimate goal of improving early warning systems that could save both lives and resources.

Yet, despite the advancement of sophisticated models, accurately predicting these occurrences is not currently within reach. Amid this complexity, the field of optics offers unique advantages for studying extreme events [2–4]. Optical systems allow researchers to recreate chaotic environments in highly controlled settings. With precise detection techniques and the ability to collect large amounts of data in short time periods, optical systems provide an ideal platform for investigating the underlying mechanisms of extreme events, and to understand how they form and evolve.

In this thesis, we focus on one such optical system: the optical fiber ring resonator. This system has become a widely used tool in nonlinear optics for exploring complex behaviors. In this setup, a continuous wave of light is injected into a passive resonator, creating a feedback loop that produces a variety of nonlinear phenomena. One of the most intriguing phenomena that arises in this setup is spatiotemporal chaos—chaotic behavior that unfolds not just over time but also across space. The complex interplay between these two dimensions results in a system that is highly sensitive to initial conditions and prone to generating extreme events. The dynamics produced by the optical fiber ring resonator are particularly challenging to predict due to their extensive chaotic nature. As the spatial and temporal scales of the chaotic dynamics increase, the complexity of the system's behavior grows exponentially. This extensive nature presents a significant challenge for long-term forecasting, as traditional models struggle to account for the vast amount of information and interactions that take place across both time and space. The chaotic signals generated by this system exhibit a mix of short- and long-range correlations, making it an ideal testbed for studying extreme events and advancing forecasting techniques.

Recent advancements in Machine Learning have opened new possibilities for forecasting in such complex systems [5–8]. Machine Learning algorithms excel in identifying patterns in large datasets offering a powerful tool for forecasting previously unpredictable occurrences. In chaotic systems like the optical fiber ring resonator, Machine Learning can find hidden correlations that are exploited to perform model-free forecasting of spatiotemporal chaos. These dynamics are common in many real-world systems, from weather patterns to fluid turbulence and optical systems. However, while promising, many Machine Learning strategies are designed to model full dynamical systems, which can be impractical for large, partially observed systems. The extensive nature of these systems means that attempting to forecast their full behavior is both computationally expensive and difficult to achieve with accuracy. Additionally, any chaotic system has a natural prediction horizon—a limit beyond which predictions become unreliable [9–11]. As the dynamics of the system evolve, forecasting beyond this horizon becomes exponentially more challenging, requiring novel approaches to extend predictive capabilities. To overcome these challenges, this work combines Machine Learning techniques with methods from information theory, particularly transfer entropy, to improve the accuracy and scope of forecasts. Transfer entropy is used to measure the directional flow of information between different parts of the system, helping to identify key precursors to extreme events. By applying transfer entropy alongside Machine Learning, we can focus on local regions of the system, isolating the critical interactions that lead to extreme events. This hybrid approach enables us to perform more targeted and reliable forecasts, even in the face of the system's complexity and chaotic nature.

A concrete real-world example of this approach can be seen in our study of wind dynamics. Using data from meteorological stations, we applied transfer entropy to analyze the flow of information between wind speed and direction across different locations. This allowed us to identify how changes in wind patterns at one station influenced conditions at another, providing insight into the propagation of extreme weather events. The application of transfer entropy to wind dynamics demonstrates the power of information theory in enhancing our understanding of chaotic, real-world systems, and highlights the broader applicability of this technique beyond optical systems.

The objective of this thesis is to explore and improve the forecasting of extreme events in chaotic systems, with a particular focus on optical systems. Through the combination of machine learning techniques and information-theoretic methods such as transfer entropy, the thesis aims to extend the prediction horizon of chaotic systems. This work contributes to both fundamental research in the field of nonlinear optics and practical applications, such as predicting real-world phenomena like wind dynamics and extreme weather events.

This thesis is structured as follows:

— In the first chapter, we begin by providing a comprehensive exploration of key concepts in dynamical systems and chaos theory. The chapter also introduces spatiotemporal chaos, which is central to the rest of this thesis. Additionally, the chapter discusses key concepts from machine learning, such as neural networks (Long Short-Term Memory networks (LSTM) and Gated Recurrent Units (GRU)), which are employed later in forecasting chaotic dynamics. Finally, the chapter introduces information theory, particularly transfer entropy, as a powerful tool for studying directional information flow in chaotic systems, laying the groundwork for the hybrid techniques developed in this work.

— In the second chapter, the focus is on forecasting the complex spatiotemporal chaos generated in an optical fiber ring resonator. The objective of the chapter is to develop machine learning models capable of predicting the full chaotic dynamics of the system. While the initial approach aims to forecast the complete system behavior, it becomes evident that the extensive nature and high dimensionality of the system pose significant challenges for long-term predictions. To address this, we introduce the use of information theory, specifically transfer entropy, to identify early-stage precursors of extreme events. This enables a shift toward local forecasting, where key regions of the system are analyzed more effectively, improving prediction accuracy and allowing for a more focused approach to managing the chaotic dynamics.

— The third chapter aims to explore the transition from soliton-based spatiotemporal chaos to turbulence in nonlinear optical systems operating in the bistable regime. Beginning with a soliton as the initial condition, the system evolves through increasingly complex dynamics as the pump parameter is raised. The primary focus is to understand how these chaotic behaviors emerge and evolve, particularly by characterizing the turbulence observed in both the phase and amplitude components of the system. Various analytical tools, including correlation functions, structure functions, and dispersion relations, are employed to study these dynamics in detail. The ultimate objective is to provide a comprehensive framework for understanding the mechanisms driving the chaotic and turbulent behaviors in this optical system, laying the groundwork for deeper insights into extreme events and their formation.

— The fourth chapter applies the tools developed in earlier chapters to a real-world system—wind dynamics. Using transfer entropy, the chapter investigates how information flows between different meteorological stations to forecast wind speed and direction. This analysis showcases the broader applicability of the techniques initially developed for optical systems and highlights the ability of transfer entropy to enhance forecasting in real-world chaotic systems like atmospheric dynamics.

— In the fifth chapter, we address one of the central challenges of forecasting chaotic systems—the predictability horizon. This chapter investigates how far into the future we can predict the behavior of chaotic systems before the forecasts become unreliable. Using the Nonlinear Local Lyapunov Exponent (NLLE) method, we estimate the maximum predictability time for chaotic systems such as the Lorenz63 model. The results contribute to a better understanding of the limitations of forecasting chaotic systems and provide insights into extending the prediction horizon in practical applications.

Chapter 1

Introduction to Dynamical Systems and Chaos Theory

Outline of the current chapter

In this chapter, we embark on a comprehensive exploration of several fundamental and interrelated fields that form the cornerstone of our research. We start by exploring the fields of dynamical systems and chaos theory, uncovering the complex behaviors and patterns that are associated. Through historical context and detailed examples, we illustrate the profound impact of chaotic dynamics on various scientific disciplines. We then introduce the concept of spatiotemporal chaos and its significance in understanding complex phenomena. We then transition to the fascinating world of Optical Frequency Combs and the Lugiato-Lefever model, highlighting their roles in nonlinear optics and the emergence of extreme events. Finally, we integrate modern approaches by introducing Machine Learning techniques and Information Theory, showcasing their potential to improve our comprehension and forecasting of chaotic systems. This chapter sets the stage for a detailed exploration of these interconnected fields, laying the groundwork for the novel contributions and insights presented in the subsequent chapters.

1.1 Dynamical Systems and Chaos theory

1.1.1 Dynamical Systems

A Dynamical System (DS) is defined by a set of states, along with rules that describe how these states evolve over time. This temporal evolution can be represented by differential equations, which model changes occurring smoothly over time, in the case of continuous-time systems. On the other hand, discrete-time DS use difference equations to model processes that change in separate steps, which can represent either natural phenomena that inherently occur at specific intervals or a methodological choice for analyzing a continuous system at distinct time points. For example, a DS can describe the population growth of a species, the oscillation of a pendulum, or the spread of a disease. The evolution of a DS is governed by a fixed rule that defines what future states follow from the current state, which can be deterministic or stochastic. In a deterministic system, a specific future state can be precisely predicted given the current state within a particular time frame. Stochastic systems, conversely, have a level of randomness where future states can only be predicted probabilistically. Some nonlinear DS exhibit chaotic behavior, which refers to deterministic yet irregular and unpredictable dynamics.

These systems can be classified as linear or nonlinear:

- **Linear Systems:** For instance, the damped harmonic oscillator, a paradigm of linear systems, can be described by the equation:

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = 0, \tag{1.1}$$

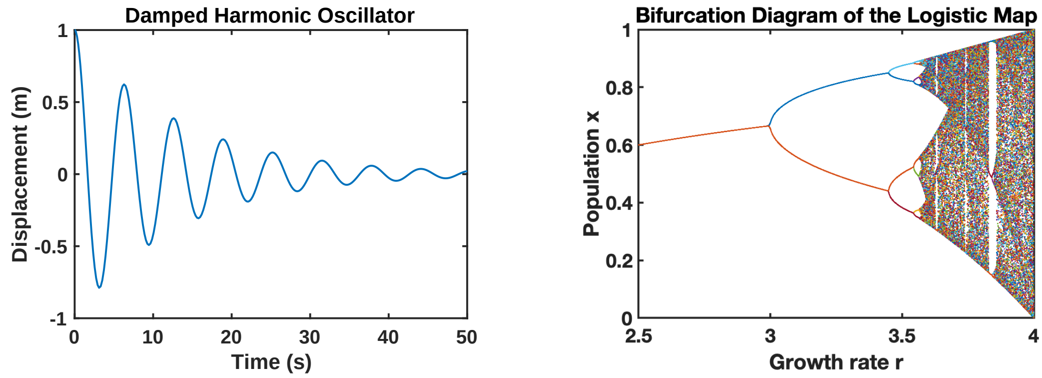
  where γ is the damping coefficient, and ω_0 is the natural frequency of the oscillator [12]. Fig. 1.1a illustrates the decay of amplitude over time in such a system.

- **Nonlinear Systems:** A famous example of a nonlinear system is the logistic map, a discrete-time population model given by:

$$x_{n+1} = rx_n(1 - x_n), \tag{1.2}$$

  which for certain values of r exhibits chaotic behavior [13]. Fig. 1.1b shows the bifurcation diagram of the logistic map, illustrating how changes in the parameter r can lead to chaotic behavior.

Nonlinear Dynamical System are particularly interesting as they can exhibit complex behaviors such as chaos, even when the underlying rules are simple and deterministic. The historical roots of DS can be traced back to the 17th century with the seminal work of Sir Isaac Newton. Newton formulated the laws of



(a) (b)

Figure 1.1: Figures demonstrating classical examples of DS. (a) The damped harmonic oscillator is a linear system where energy dissipates over time. (b) Bifurcation diagram of the logistic map, a nonlinear system, demonstrating the transition from stability to chaos as the parameter r increases, showing complex dynamical behavior including periodic doubling leading to chaos.

motion and universal gravitation, which could be used to predict the behavior of celestial bodies with remarkable precision. His differential equations describing the two-body problem such as the Earth orbiting the Sun laid the groundwork for classical mechanics [14]. However, the more complex three-body problem, involving the gravitational interaction between three celestial bodies, remained unsolved and would puzzle mathematicians for centuries [15].

In the late 19th century, Henri Poincaré developed qualitative methods to study the behavior of DS, particularly in the context of the three-body problem. Poincaré's innovative approaches laid the groundwork for the modern theory of DS, including the fields of topology and bifurcation theory. He was the first to recognize that deterministic systems could exhibit aperiodic behavior, a notion that would later evolve into the concept of chaos [16]. In the 20th century, the development of technologies such as radio and radar illustrated the practical applications of nonlinear oscillators, a key concept within DS. These applications underscored the importance of understanding the behavior of systems described by nonlinear differential equations [17]. The advent of computers in the second half of the 20th century marked a turning point in the study of DS. The ability to perform numerical simulations allowed scientists to visualize and explore the complex behaviors that Poincaré had only imagined. Edward Lorenz's 1963 discovery of the chaotic behavior of a simplified model of atmospheric convection demonstrated the profound implications of sensitivity to initial conditions and brought the study of chaos to the forefront of scientific inquiry [1].

Today, the field of DS encompasses a vast array of phenomena across various disciplines, from the predictability of planetary motion to the complexities of weather forecasting and the rhythms of biological systems. The theory continues to provide a fundamental framework for understanding both the deterministic nature and inherent unpredictability of many natural and engineered systems [18].

1.1.2 Chaos in Dynamical Systems

Chaos theory explores the behavior of Dynamical System that are highly sensitive to initial conditions—a phenomenon popularly known as the butterfly effect. This sensitivity means that small differences in the initial setup of a system can lead to vastly divergent outcomes, making long-term prediction difficult or impossible, even in systems that are deterministic in nature. Chaos is defined as the apparent randomness that emerges from simple, deterministic systems due to their sensitivity to initial conditions. Unlike truly random behavior, which is inherently unpredictable and not governed by deterministic laws, chaotic behavior arises from well-defined mathematical models and rules. The unpredictability in chaotic systems stems from our inability to measure initial conditions with infinite precision. The primary characteristic of chaotic systems is their unpredictability over long periods due to the exponential growth of errors in initial condition measurements. Despite this unpredictability, chaotic systems are not without order; they often display self-similar patterns across different scales of observation. Chaos theory has profound implications across various scientific disciplines, including meteorology, engineering, economics, biology, and more. Its discovery has led to a better understanding of the limits of prediction and control in complex systems, reshaping our approach to modeling and analysis in science and engineering.

1.1.3 Brief History of Chaos

Chaos theory, as we understand it today, has evolved through a history of remarkable discoveries that have reshaped our understanding of DS. As we mentioned in Section 1.1.1, Henri Poincaré laid the groundwork in the late 19th century when he explored solutions to the three-body problem and recognized the potential for irregular and aperiodic orbits [19]. These insights hinted at the complex behavior inherent in DS, which could not be explained by traditional Newtonian mechanics. This notion of irregular, complex behavior in systems governed by deterministic laws was later addressed in the seminal work of Edward Lorenz in 1963 [1], who, through his reduced models of weather patterns, laid the foundation for what we understand as chaos today. It was not until Steve Smale, in 1967, that the geometrical underpinnings of chaos started to be unveiled, showing how these systems could behave in a complex yet deterministic fashion [20]. This was further confirmed by Ruelle and Takens' challenge to Landau's theory on turbulence, suggesting that even simple systems of coupled oscillators could exhibit chaotic behavior [21]. The 1970s and 1980s saw a flourish of experimental observations confirming chaotic behavior in various systems. Notably, Golub and Swinney observed chaos in a fluid dynamics experiment [22], and Li and Yorke's mathematical findings suggested that even periodic systems could have chaotic underpinnings, famously captured in the phrase "period three implies chaos" [23]. By the 1980s, the field had expanded dramatically, with chaos being observed in systems as diverse as electronics, chemistry, and optics. Entering the late 20th century, the focus shifted to the control and synchronization of chaos, with notable contributions from Ott, Grebogi, and Yorke in 1990, opening doors to new methods of managing systems that were once thought uncontrollable [24]. Chaos theory continues to be a vibrant field of study, with applications ranging from forecasting weather to modeling economic systems. The discovery that even simple, deterministic systems can behave unpredictably has profound implications for how we understand and interact with the world around us.

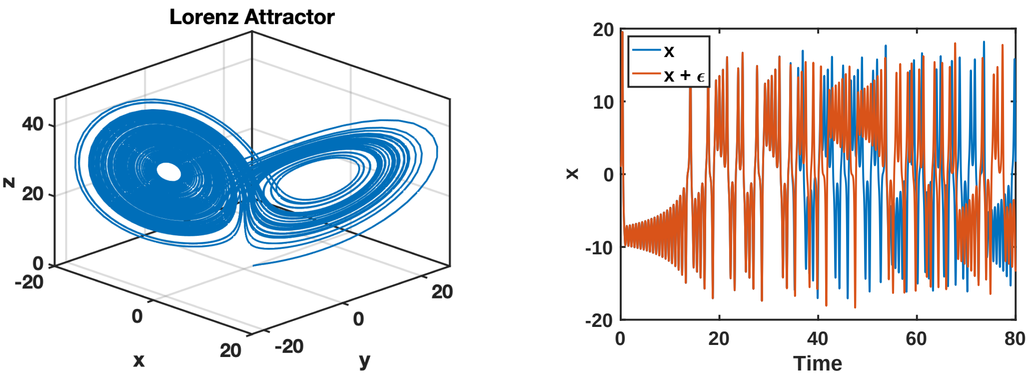
1.1.4 Lorenz and Rössler Systems: A Paradigm of Chaos

Lorenz system

The Lorenz system, uncovered by Edward Lorenz in 1963 during his exploration of atmospheric convection, stands as a seminal example within chaos theory. This model, through its set of three simple differential equations, reveals the profound complexity underlying chaotic systems. The equations govern the movement of fluid within a layer that is heated from below and cooled from above, manifesting the unpredictability inherent in such systems [1]. The essence of the Lorenz system lies in its demonstration of sensitive dependence on initial conditions, popularly known as the butterfly effect. This principle proves that even minuscule numerical differences in the starting conditions of a system can lead to vastly divergent outcomes, rendering long-term prediction a formidable challenge. The Lorenz attractor, a complex, butterfly-shaped structure emerging from the system's equations, visually encapsulates this concept, symbolizing the intricate dynamics at play within chaotic phenomena [25, 26]. The attractor's fractal nature and its representation of the system's phase space dynamics are symbolic of chaos, underscoring the unpredictable behavior that can arise from deterministic rules. The governing equations of the Lorenz system are short yet profound:

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \beta z, \end{cases} \tag{1.3}$$

where σ, ρ, and β represent the system's parameters, linked to the Prandtl number [27], Rayleigh number [28], and specific physical dimensions of the convection box, respectively. $\dot{x}$ denotes the time derivative of x. Fig. 1.2a is an illustrative example of the Lorenz attractor, showcasing the complex, butterfly-shaped structure that symbolizes the essence of chaos theory. Fig. 1.2b compares the time evolution of the x variable under two different initial conditions, separated by a small error of $\epsilon = 10^{-9}$. This comparison vividly highlights how slight variations in starting conditions can dramatically influence the system's trajectory over time, providing a powerful visual representation of chaos theory in action. Because of its simplicity, the Lorenz system has become one of the most extensively studied examples of nonlinear dynamics. Its ability to generate chaotic attractors from a straightforward set of equations continues to intrigue and challenge scientists, offering deep insights into the unpredictable yet structured world of chaotic systems.



(a) (b)

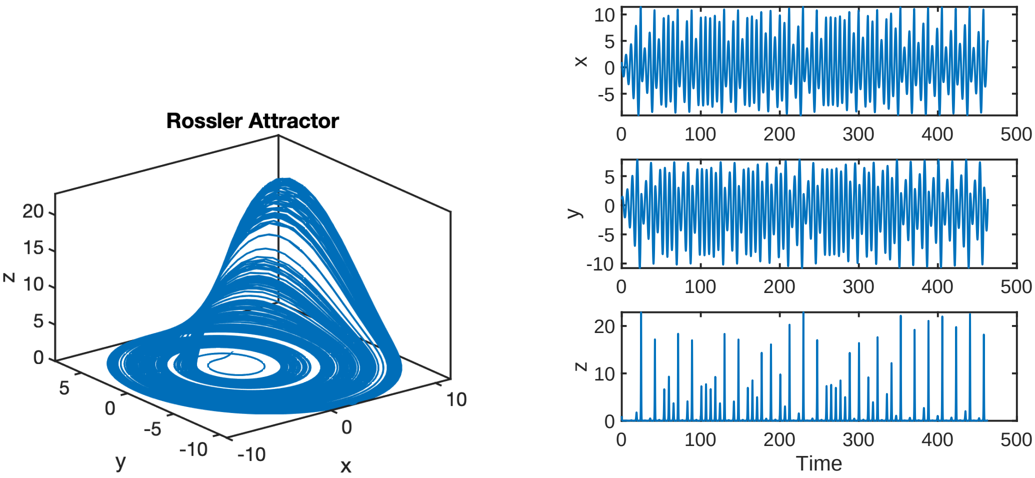
Figure 1.2: (a)Visualization of the Lorenz attractor, demonstrating the complex dynamics of the system. This figure highlights the butterfly-shaped trajectory that is characteristic of chaotic behavior, emphasizing the sensitive dependence on initial conditions and the unpredictability of the system's long-term behavior. (b) Time evolution of the x variable in the Lorenz system under two different initial conditions, separated by a very small value of 10^{-9}. This comparison illustrates how minor discrepancies in starting conditions can lead to significant divergences in the system's trajectory over time.

Rössler system

In 1976, Otto Rössler presented a system related to the Lorenz63 system [29], explicitly designed to produce continuous temporal chaos [30]. While Rössler wrote down several systems that would be able to create continuous chaos, the most commonly used is the system given by [31]:

$$\begin{cases} \dot{x} = -y - z, \\ \dot{y} = x + ay, \\ \dot{z} = b + z(x - c), \end{cases} \tag{1.4}$$

where the parameters a, b, c determine whether the system evolves periodically, chaotically, or converges to a static solution. For $a = 0.5$, $b = 2.0$, and $c = 4.0$, the system evolves chaotically [31]. The attractor for this parameter choice is shown in Fig. 1.3, along with the individual x, y, and z components plotted against time.



(a) (b)

Figure 1.3: (a) Visualization of the Rössler attractor in three-dimensional phase space, illustrating the chaotic trajectory that emerges under specific system parameters $a = 0.5$, $b = 2.0$, and $c = 4.0$. The complex, swirling dynamics highlight the system's sensitive dependence on initial conditions. (b) Time evolution of the Rössler system's variables: x (top), y (middle), and z (bottom). These plots demonstrate the non-periodic behavior characteristic of chaos, with each variable exhibiting erratic and unpredictable fluctuations over time.

1.1.5 Lyapunov Exponents

The concept of Lyapunov Exponent (LE) is central to the study of chaotic systems, offering a quantitative measure of the sensitive dependence on initial conditions. LE quantify the average exponential rate at which trajectories in a Dynamical System diverge from one another. Positive LE are indicative of chaos, suggesting that even slight differences in the system's initial state can result in exponential separation of trajectories over time [32], as demonstrated in Fig. 1.4. In contrast, negative exponents signify convergence towards stable states such as fixed points or periodic orbits, while a zero value indicates neutral stability often observed in limit cycles [33].

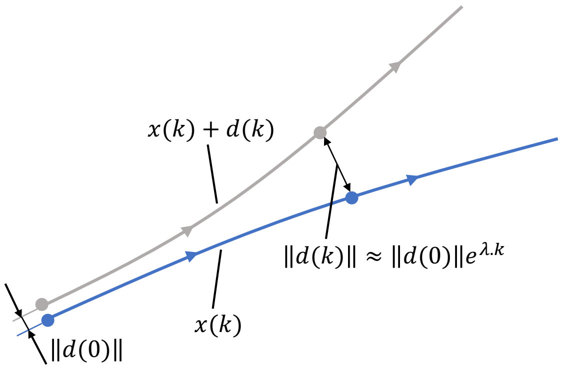


Figure 1.4: Illustration of two trajectories in phase space that initially start close together and exponentially diverge, exemplifying the concept of LE in chaotic systems. This figure emphasizes the exponential growth of small perturbations, highlighting the unpredictable and sensitive nature of chaotic systems.

The largest Lyapunov Exponent, denoted by $\lambda_{\max}$, is computed by reconstructing the state space through techniques like Takens' embedding theorem and then assessing the rate of trajectory divergence. This calculation is essential for understanding the dynamics of chaotic systems and is visualized through the exponential divergence of two closely started trajectories in phase space [34]. The formula for $\lambda_{\max}$ is given by:

$$\lambda_{\max} = \frac{1}{k\Delta n(S-1)} \sum_{n=1}^{S-k} \log \frac{d(k)}{d(0)} \tag{1.5}$$

where $d(0)$ is the initial separation between trajectories, $d(k)$ is the separation after k discrete time steps, and S is the total number of data points [35]. Table 1.1 encapsulates the relationship between $\lambda_{\max}$ and the nature of a system's dynamics, distinguishing between fixed points, limit cycles, chaotic behavior, and noise-like states. In summary, LE serves as a fundamental tool for identifying and characterizing chaos within complex systems.

Dynamics	**Largest LE**
Fixed Point	$\lambda_{\max} < 0$
Limit Cycle	$\lambda_{\max} = 0$
Chaos	$0 < \lambda_{\max} < \infty$
Noise	$\lambda_{\max} \to \infty$

Table 1.1: Largest LE for different dynamical behaviors.

1.1.6 Spatiotemporal Chaos

In the previous sections, we explored the complexities of temporal chaos, which is characterized by its sensitivity to initial conditions over time within a system's trajectory. This type of chaos, exemplified by the Lorenz system and described through LE, highlights the unpredictable yet deterministic nature of dynamic systems observed at a specific point in space. We will now extend this conversation to Spatiotemporal Chaos, which incorporates the spatial dimension into chaotic analysis. This broader type of chaos is not limited to just temporal evolution but is present across large spatial areas, characterized by a continuous spectrum of positive LE (which we will detail later in Section 1.3.6), as opposed to the discrete values found in purely temporal chaos.

Spatiotemporal Chaos provides a crucial framework for understanding complex phenomena where the interaction between spatial and temporal elements is intricate, resulting in patterns that emerge from the collective dynamics of spatially extended systems. Such systems are pivotal in a variety of contexts, from predicting weather patterns to understanding the spread of diseases and managing traffic flows. The foundation of STC was significantly advanced by M.C. Cross and P.C. Hohenberg in the late 20th century through their seminal work on pattern formation in systems far from equilibrium, highlighting the complex behaviors arising from spatial interactions and disturbances [36]. Additional developments by researchers like Y. Kuramoto and G. Nicolis have further explored how coupled oscillatory fields and reaction-diffusion systems under non-equilibrium conditions can exhibit chaotic behavior, profoundly affecting the formation and evolution of patterns across large spatial scales [37, 38].

These studies have opened new avenues in physics and applied mathematics, significantly impacting diverse fields such as meteorology, ecology, and social sciences by enhancing our understanding of the spatial distribution and temporal dynamics of complex systems. Through the collective efforts of these researchers, STC has become an indispensable concept in the study of complex systems, aiding in the prediction and management of phenomena that exhibit unpredictable behavior across both space and time.

Examples of Spatiotemporal Chaos

As we continue our exploration of STC, it is crucial to investigate specific systems where this complex phenomenon is prominently displayed. STC, often emerging in systems governed by partial differential equations, presents unique challenges compared to temporal chaos, which can be modeled more straightforwardly with Ordinary Differential Equations (ODE). Among the notable systems that exhibit STC, the Complex Ginzburg-Landau Equation (CGLE) and the Kuramoto-Sivashinsky Equation (KSE) are particularly significant. Each of these equations encapsulates distinct mechanisms and effects of STC, providing deep insights across a wide range of applications. These models serve as fundamental tools for understanding the intricate dynamics that characterize STC in various scientific fields.

The CGLE is crucial in the study of nonlinear waves and pattern formation in nonequilibrium systems. It describes a wide range of phenomena, including superconductivity, superfluidity, Bose-Einstein condensates, and nonlinear wave dynamics in fluids [39]. The CGLE is given by:

$$\frac{\partial A}{\partial t} = A + (1 + i\alpha)\nabla^2 A - (1 + i\beta)|A|^2 A, \tag{1.6}$$

where A represents the complex amplitude of the wave, and α and β are real parameters that modulate the nonlinear interaction and dispersion, respectively. This equation illustrates how complex interactions and turbulence within wave patterns can evolve, demonstrating STC in a mathematically elegant form.

The KSE on the other hand, captures dynamics in systems experiencing instabilities leading to turbulence. Originally derived to study the instability in laminar flames, the KSE is applicable to various hydrodynamic systems, chemical reactions, and heat transfer phenomena [40]. The KSE is formulated as:

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4}, \tag{1.7}$$

where u represents the state variable, such as temperature in a thermal convection layer or concentration in a chemical reaction. This equation is renowned for its chaotic solutions, providing a window into the behavior of systems far from equilibrium.

Both the CGLE and KSE exemplify the critical nature of STC in understanding the dynamics of various scientific and engineering systems. These models demonstrate the intricate patterns that emerge from the nonlinear dynamics and are instrumental in developing theoretical and computational tools to analyze chaos over extensive spatial and temporal scales [36]. By studying these equations and their implications, researchers continue to unravel the complexities of STC, enhancing our understanding of chaotic behavior in natural and engineered systems. Insights from such studies are invaluable, broadening our capabilities in fields ranging from meteorology and environmental science to physics and beyond.

1.2 Optical Frequency Combs

Expanding upon the basic knowledge of chaotic and Dynamical System discussed in previous parts, we now turn our attention to an especially intriguing use in the field of spectroscopy and precise measurement: Optical Frequency Combs (OFC). Over the past few decades, these combs, composed of a series of phase-locked, equally spaced laser frequency lines, have received significant interest due to their precision and the broad spectrum of their applications. These combs serve as highly accurate rulers, finding diverse applications ranging from metrology [41, 42] with Lidars [43] to astrophysical spectrometers for detecting Earth-like exoplanets [44, 45], as well as in ultra-precise spectroscopy and molecular fingerprinting [46–48]. The 2005 Nobel Prize in Physics awarded to John L. Hall and Theodor W. Hänsch highlighted the global significance of frequency combs, particularly for their use in precision hydrogen spectroscopy, influencing fundamental research and real-world metrology applications.

The development of OFC is closely associated to advancements in laser technology, particularly through mode-locked lasers that emit a rapid succession of ultra-short light pulses. Each pulse carries a series of phase-coherent optical frequency lines that come together to form the comb's structure. As depicted in Fig. 1.5, the time domain representation of mode-locked frequency combs shows a periodic train of pulses with a repetition rate f_{rep} which corresponds to an equidistant spectrum in the frequency domain. This visual representation helps illustrate how the spacing of the lines, given by f_{rep}, and the carrier-envelope phase shift $\Delta\phi_0$ influence the structure and stability of the frequency comb. The figure effectively demonstrates the interplay between time and frequency domains that is central to the functionality of OFC.

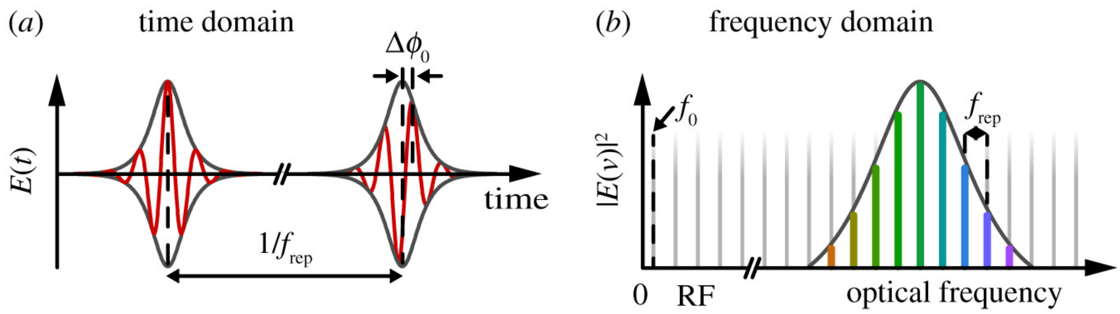


Figure 1.5: Time and frequency domain picture of mode-locked frequency combs. A periodic train of pulses with a repetition rate f_{rep} (a) corresponds to an optical frequency comb spectrum of equidistant lines in the frequency domain (b). The line spacing is given by f_{rep}. The offset f_0 of the frequency comb spectrum relates to the carrier-envelope phase shift $\Delta\phi_0$ between two consecutive pulses via $f_0 = f_{rep} \cdot \Delta\phi_0/2\pi$. Figure taken from [49].

Beyond traditional mode-locking techniques [50], other methods have also been developed to generate these OFC, including parametric four-wave mixing in highly nonlinear fibers [51], spontaneous four-wave mixing in quantum cascade lasers [52], and difference frequency generation [53].

Additionally, the generation of Kerr combs in microresonators has emerged as a compact and efficient alternative [54], exploiting the nonlinear optical properties of these systems to generate broad and stable combs across a wide spectral range.

Microresonators, especially those with high-quality factors, have proven to be optimal for generating OFC due to their ability to efficiently confine light [55–57] and their low energy requirements. The use of Continuous-Wave lasers to pump these microresonators facilitates the generation of Kerr combs through enhanced four-wave mixing, driven by the intrinsic nonlinearity of the medium, marking a significant advancement in the field. Fig. 1.6 illustrates a typical microresonator setup used in the creation of OFC.

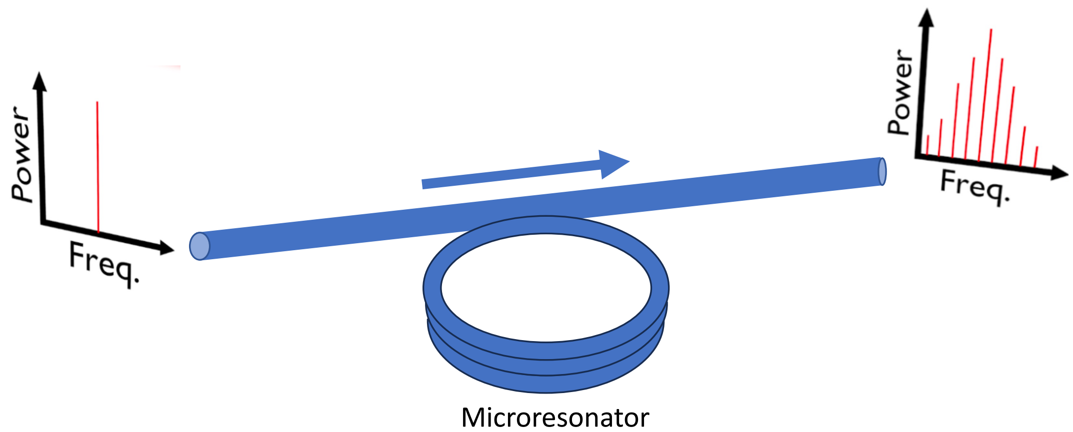


Figure 1.6: Schematic diagram of a microresonator used in generating OFC.

In our work, we consider a passive Kerr resonator made of an optical fiber ring, which we will talk about more in Chapter 2.

1.3 Lugiato-Lefever Model

The Lugiato-Lefever Equation (LLE) was first introduced by Luigi Lugiato and René Lefever in 1987 to describe spatial dissipative structures in nonlinear optical systems, specifically within passive optical resonators. This model has since been pivotal in understanding various phenomena in nonlinear optics, such as Kerr Frequency Comb (KFC), optical solitons, and Modulation Instability (MI) [58–60].

By coupling the ends of an optical fiber, the propagation of light in the resulting passive resonator can be modeled by the Nonlinear Schrödinger Equation

(NLSE), augmented with boundary conditions or Ikeda map, as follows [61–63]:

$$\frac{\partial A^{(j)}(z,T)}{\partial z} = -\frac{\alpha_f}{2}A^{(j)}(z,T) - i\frac{\beta_2}{2}\frac{\partial^2 A^{(j)}(z,T)}{\partial T^2} + i\gamma|A^{(j)}(z,T)|^2 A^{(j)}(z,T), \tag{1.8a}$$

$$A^{(j+1)}(0,T) = \sqrt{\theta}E_i(T) + \sqrt{\rho}A^{(j)}(L,T)e^{-i\Phi_0}, \tag{1.8b}$$

where $A^{(j)}(z,T)$ is the envelope of the electric field that circulates within the cavity at the j^{th} round trip, α_f the absorption of the fiber, L the cavity length, γ the nonlinear Kerr coefficient of the fiber, β_2 the group-velocity dispersion, Φ_0 is the linear phase shift, θ and ρ are the transmission and reflection coefficient respectively. The independent variable z refers to the longitudinal coordinate, while T is the time in a reference frame moving with the group velocity of the light. From the analytical point of view, except for the Linear Stability Analysis (LSA), it is a hard task to study the nonlinear evolution of such a system directly via Eqs. 1.8. For large enough cavity finesse, the mean-field approximation is often used to describe the full map equations. Consequently, after one roundtrip, the solution of Eq. 1.8a can be written in the form:

$$\begin{aligned} A^{(j)}(L,T) &\approx A^{(j)}(0,T) + L\left.\frac{\partial A^{(j)}(0,T)}{\partial Z}\right|_{Z=0} \\ &\approx \left[1 - \frac{\alpha_f L}{2} - i\frac{\beta_2 L}{2}\frac{\partial^2}{\partial T^2} + i\gamma L\left|A^{(j)}(0,T)\right|^2\right]A^{(j)}(0,T). \end{aligned} \tag{1.9}$$

For resonant pumping, the acquired phase after a roundtrip is $2k\pi$ where the integer k labels the cavity resonances. The distance to the closest cavity resonance is measured by introducing $\delta = 2k\pi - \Phi_0$. Assuming $\delta \sim \theta \ll 1$, the term $\sqrt{\rho}\exp(-i\Phi_0)$ in Eq. 1.8b can be written in the form:

$$\sqrt{\rho}\exp(-i\Phi_0) = \sqrt{1-\theta}\exp(-i\Phi_0) \approx (1-\frac{\theta}{2})(1-i\delta) \approx 1 - \frac{\theta}{2} - i\delta. \tag{1.10}$$

Using this approximation after inserting Eq. 1.9 in Eq. 1.8b leads to:

$$A^{(j+1)}(0,T) \approx \sqrt{\delta}E_i(T) + \left[1 - \frac{\theta}{2} - \frac{\alpha L}{2} - i\delta - i\frac{\beta_2 L}{2}\frac{\partial^2}{\partial T^2} + i\gamma L|A^{(j)}(0,T)|^2\right]A^{(j)}(0,T).$$

From this equation, one can easily identify the quantity that measures the change in the temporal profile from one roundtrip to another at the coupler as $A^{(j+1)}(0,T) - A^{(j)}(0,T) \approx t_R \partial_t' A(t',T)$ with $t' = jt_R$. Finally setting:

$$S = E_i\sqrt{\gamma\theta L/\alpha^3} \tag{1.11a}$$
$$\psi = A\sqrt{\gamma L/\alpha}, \tag{1.11b}$$
$$t = \frac{\alpha t'}{t_R}, \tag{1.11c}$$
$$\tau = T\sqrt{\frac{2\alpha}{\beta_2 L}}, \tag{1.11d}$$
$$\alpha = \frac{\theta + \alpha_f L}{2}, \tag{1.11e}$$
$$\eta = \text{sign}(\beta_2) = \pm 1, \tag{1.11f}$$
$$\Delta = \frac{\delta}{\alpha} = \frac{2k\pi - \Phi_0}{\alpha}, \tag{1.11g}$$

the mean-field evolution is given by the LLE [58, 62]:

$$\frac{\partial\psi(t,\tau)}{\partial t} = S - (1 + i\Delta)\psi - i\eta\frac{\partial^2\psi}{\partial\tau^2} + i\,|\,\psi\,|^2\,\psi \tag{1.12}$$

Here $\psi(t,\tau)$ represents the normalized slowly varying envelope of the electric field circulating within the cavity, S denotes the pump strength, Δ is the frequency detuning, t corresponds to the slow evolution of ψ over successive round-trips, τ accounts for the fast dynamics that describe how the electric field envelope changes along the fiber[1]. For the normal dispersion regime, the group-velocity dispersion $\beta_2 > 0$ and $\eta = 1$, and for the anomalous dispersion regime $\beta_2 < 0$ and $\eta = -1$. The LLE is foundational in the study of nonlinear dynamics within optical systems. This equation was originally developed to describe spatial dissipative structures in passive optical systems, where the interplay between diffraction, nonlinearity, and cavity detuning results in complex behaviors such as solitons, periodic patterns, and chaos [58, 59, 62]. Note

[1]The LLE (Eq.1.12) is equivalent to a 1D system. In the following, when we refer to a spatiotemporal system, "time" will denote the round-trip time t, while "space" will correspond to the fast time τ.

that, before laser systems, Eq. 1.12 has been derived in early reports to describe the plasma driven by radio frequency field [64, 65] and the condensate in the presence of an applied AC field [66].

1.3.1 Steady states

In the homogeneous ($\frac{\partial^2 \psi_S}{\partial \tau^2} = 0$) and steady-state ($\frac{\partial \psi_S}{\partial t} = 0$) case, the LLE simplifies to:

$$0 = -(1 + i\Delta)\psi + i|\psi|^2\psi + S \tag{1.13}$$

Rewriting this equation in terms of real and imaginary parts, let $\psi = \psi_r + i\psi_i$ and assuming S is real ($S_i = 0$), we get two equations:

$$\begin{cases} S = \psi_r - \Delta\psi_i + |\psi|^2\psi_i \\ 0 = \psi_i + \Delta\psi_r - |\psi|^2\psi_r \end{cases}$$

By solving these equations for the steady-state values, we find:

$$\psi_i = \psi_r(I_s - \Delta) \tag{1.14}$$

where $I_s = |\psi_s|^2$ is the intracavity field intensity. Rewriting the steady-state field ψ_s in terms of amplitude $\sqrt{I_s}$ and phase ϕ_s:

$$\psi_s = \sqrt{I_s}e^{i\phi_s} \tag{1.15}$$

Thus, the real and imaginary parts are: $\psi_r = \sqrt{I_s}\cos(\phi_s)$ and $\psi_i = \sqrt{I_s}\sin(\phi_s)$. Using Eq. 1.14, we can calculate the steady state phase:

$$\phi_s = \tan^{-1}(I_s - \Delta) \tag{1.16}$$

The homogeneous and steady state solutions of Eq. 1.12 obey the cubic equation:

$$|S|^2 = I_s[1 + (I_s - \Delta)^2] \tag{1.17}$$

It is well known that this equation has one, two, or three real-valued solutions depending on the parameters Δ and S. Multiple solutions may arise in a polynomial equation when it has local extrema; in our case, this condition requires the existence of critical values of I_s such that the partial derivative is null:

$$\frac{\partial |S|^2}{\partial I_s} = 3I_S^2 - 4\Delta I_s + \Delta^2 + 1 = 0 \tag{1.18}$$

This condition yields a quadratic equation with a discriminant equal to $4(\Delta^2 - 3)$; therefore, if $\Delta < \sqrt{3}$, there are no such critical values for I_s whereas, for $\Delta \geq \sqrt{3}$, these critical values are:

$$I_{S,SN_{1,2}} = \frac{2\Delta}{3} \pm \frac{1}{3}\sqrt{\Delta^2 - 3}. \tag{1.19}$$

and the corresponding pumping terms are:

$$S_{\pm}^2(\Delta) = \frac{2\Delta \mp \sqrt{\Delta^2 - 3}}{3}\left[1 + \left(\frac{\sqrt{\Delta^2 - 3} \pm \Delta}{3}\right)^2\right] \tag{1.20}$$

The critical detuning is obtained by imposing $I_{S,SN_1} = I_{S,SN_2}$ which gives $\Delta_c = \sqrt{3}$.

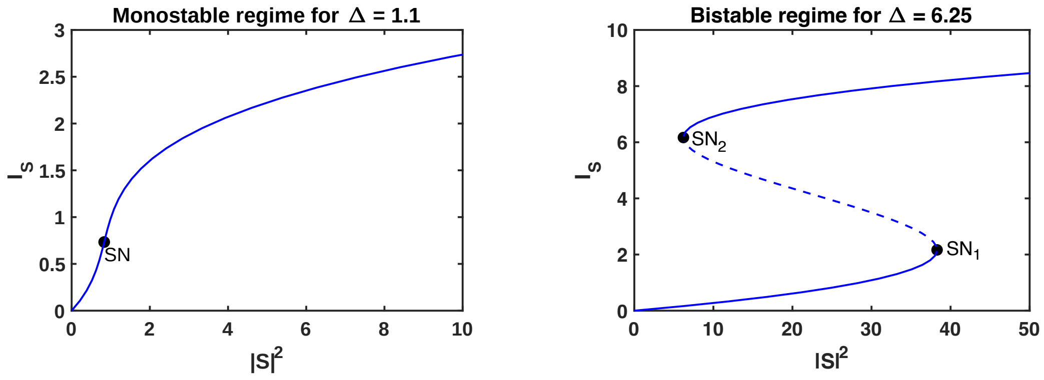
- For small detunings $\Delta < \sqrt{3}$, the HSS Eq. 1.17 has only one solution such that there is a single steady state I_S for a given pump S^2, the system is **monostable** (single-valued transmission curve).

- For detuning $\Delta > \sqrt{3}$, the HSS Eq. 1.17 has three solutions for I_S for a given value of the pump S^2 with two turning points $SN_{1,2}$. Two solutions are stable to homogeneous perturbations while the other is unstable, and thus the homogeneous solution is **bistable** (presence of a Hysteresis).

In Fig. 1.7, we can observe the monostable regime for $\Delta = 1.1$ (Fig. 1.7a) and bistable regime for $\Delta = 6.25$ (Fig. 1.7b).

1.3.2 Linear Stability Analysis

The stability analysis of the LLE reveals critical insights into the conditions under which complex behaviors such as Modulation Instability (MI) and soliton



(a) (b)

Figure 1.7: In figure (a) we show the HSS in the monostable regime for $\Delta = 1.1$ and in figure (b) an example of the HSS solution in the bistable regime, here $\Delta = 6.25$.

formations occur. Solitons form when there is a balance between nonlinearity and dispersion, leading to stable, localized wave packets. MI arises when perturbations on a continuous wave background grow exponentially, driven by the interplay between nonlinearity and dispersion.

To analyze the stability of the homogeneous steady-state solution of the LLE, we consider perturbations of the form $\psi = \psi_s + \delta\psi$, where ψ_s is the steady-state solution and $\delta\psi$ is a small perturbation. The perturbation is assumed to have the form $\delta\psi \propto e^{\lambda t + i\omega\tau}$. Substituting this into the LLE and linearizing around ψ_s gives the dispersion relation:

$$(\lambda + 1)^2 + (\Delta - 2I_s + \omega^2)^2 - I_s^2 = 0 \tag{1.21}$$

Solving for the eigenvalues λ:

$$\lambda_\pm = -1 \pm \sqrt{I_s^2 - (\Delta - 2I_s + \omega^2)^2} \tag{1.22}$$

The eigenvalues λ_- always being negative, whatever $\omega \in \mathbb{R}$, only λ_+ is of interest for stability analysis. The system becomes unstable if λ_+ has a positive real part.

The boundary between stable and unstable regimes is given by $\lambda_+ = 0$:

$$-1 + \sqrt{I_s^2 - (\omega^2 + \Delta - 2I_s)^2} = 0 \tag{1.23}$$

The specific conditions for MI differ between the monostable and bistable regimes, with different thresholds for detuning and pump power. In the following sections, we will detail these phenomena, providing a thorough explanation of the conditions for soliton formation and MI. For more detailed calculations and a deeper mathematical treatment, please refer to Appendix A.

1.3.3 Modulation Instability

Modulation Instability, also known as Turing instability [36, 67–69], is a critical phenomenon in nonlinear optical systems. It occurs when a Continuous-Wave (CW) solution becomes unstable due to perturbations, resulting in the formation of a periodic pattern in the temporal domain and symmetrical sidebands on both sides of the pump frequency in the spectral domain. This effect was first observed in nonlinear fiber optics in 1986 by K. Tai et al. [70] in the anomalous dispersion region. MI has also been investigated in cavities, where it was first observed in 1988 by M. Nakazawa et al. [71]. In 1997, Coen et al. [61] highlighted that MI could be extended to normal dispersion region. MI in passive cavities has also been widely studied in anomalous dispersion since it operates at the early stage of higher nonlinear structures such as temporal Cavity Solitons (CS) [60], whose spectral counterpart corresponds to frequency combs [72]. The conditions for MI are influenced by the interplay between Kerr nonlinearity and the system's dispersion, and they differ significantly between the monostable and bistable regimes.

MI in Monostable Regime ($\Delta < \sqrt{3}$)

In the monostable regime, the system is characterized by a single stable steady-state solution. MI can still occur under specific conditions:

- Instability Condition: The perturbation growth rate λ_+ must be positive.

According to the dispersion relation (Eq. 1.22), the condition for MI to occur is: $I_s^2 - (\Delta - 2I_s + \omega^2)^2 > 1$.

- Critical Points: The HSS loses stability at $I_s = 1$ with $\omega_c = \sqrt{2-\Delta}$.

- Patterned Solutions: MI leads to a periodic solution in τ, which considering τ as a spatial coordinate, can be viewed as a pattern solution. This patterned solution can arise either supercritically when $\Delta < 41/30$ or subcritically when $\Delta > 41/30$.

MI in Bistable Regime ($\Delta > \sqrt{3}$)

In the bistable regime, the system has two stable steady-state solutions separated by an unstable one as we showed in Section 1.3.1. This regime supports more complex dynamics, and MI can occur under different conditions:

- For $\sqrt{3} < \Delta < 2$, MI occurs at $I_s = 1$. The critical values for MI are the two values derived from the saddle-node bifurcation analysis (Eqs. 1.19 derived in the previous section).

- For $\Delta > 2$, the system transitions directly to optical turbulence above the upper saddle-node bifurcation point SN_2, while stable periodic patterns persist below the threshold. This coexistence can lead to localized structures such as CS.

For MI to occur, the real part of the eigenvalue $\mathfrak{Re}(\lambda)$ must be positive. A necessary condition in all cases is that the CW intracavity power $I_s \geq 1$, which implies that the minimum driving power required for intracavity MI is $S \geq 1$. In the anomalous dispersion regime, MI requires $I_s > \Delta/2$ for $\Delta > 2$, making the entire upper branch of the homogeneous response unstable while the lower one remains stable [73]. Conversely, in the normal dispersion regime, MI occurs only at the end of the CW lower branch when $I_s < \Delta/2$ [74]. The MI pattern solution emerges supercritically from the homogeneous state when $\Delta < 41/30$ and subcritically when $\Delta > 41/30$ [58, 73]. In the latter case, a periodic pattern can coexist with a stable CW solution (the lower branch when the bistability of the homogeneous state is present) over a certain range of parameters. This coexistence underpins the existence of CS [75, 76], which are localized structures formed by the periodic MI pattern connected by fronts with the CW solution [77].

1.3.4 Cavity solitons

Solitons in the LLE arise when dispersion and nonlinearity balance each other, resulting in a stable, localized wave packet that maintains its shape over time. Additionally, a balance between gain and dissipation defines the soliton's amplitude. Cavity Solitons (CS) are solitons that exist in passive nonlinear resonators, where the loss compensation is achieved through coherent external forcing [78]. CS emerges from the homogeneous CW response at the CW up-switching point. Initially, they are unstable, very broad, and have a low peak power above the CW background, appearing infinitely wide as they approach the CW up-switching point. As the system parameters change, the CS branch eventually folds and becomes stable. At higher detuning values, a Hopf bifurcation occurs, causing a soliton to destabilize and evolve into a breather, that is, a soliton whose amplitude varies periodically in time [79, 80]. For even larger parameter values, the dynamics become more complex and can lead to chaotic behavior [79, 81].

1.3.5 Spatiotemporal Chaos and its two routes

It is well known that chaos can potentially arise in any nonlinear system with at least three degrees of freedom. A third condition, which is rather empirical but intuitively evident, is the requirement of strong excitation. The LLE satisfies these conditions due to its high nonlinearity and its infinite number of dimensions and therefore predicts the occurrence of STC. Practically, almost all high-dimensional and nonlinear systems exhibit chaotic behavior when strongly excited. The first unambiguous evidence of chaos in Kerr combs was provided in Ref. [82] using the computation of LE, which were found to be positive in the case of strong pumping. Other studies have also shed more light on the chaotic dynamics of Kerr combs [83]. From a more general perspective, two main routes to chaos can be identified [84].

- Route via Unstable Turing Patterns: If $\Delta < \sqrt{3}$, chaos originates from the destabilization of Turing patterns. Starting from a low-amplitude noise initial condition, as the pump S^2 is increased, a sequence of bifurcations occurs. These bifurcations lead to the Turing patterns becoming unstable and oscillating in time. Further increases in pump power drive the system into a chaotic state, characterized by rapidly fluctuating peaks of diverse amplitudes.

- Route via CS: When $\Delta > \sqrt{3}$, the system's stable structures are CS. As the pump power increases, the solitons become unstable and evolve into a breather, whose amplitude varies periodically over time. With further increases in pump power, the system transitions into chaotic behavior, the unstable solitons are in a "turbulent" regime which is characterized by the pseudorandom emergence of very sharp and powerful peaks.

A statistical analysis shows that these peaks which are of rare occurrence and very high intensity [85] qualify as rogue waves or Extreme Events (EE) [86, 87].

1.3.6 Characterization of spatiotemporal chaos

LE measure the sensitivity of the system to initial conditions, and as we discussed earlier in Section 1.1.5, a positive LE is evidence of chaos, but to distinguish between various complex behaviors such as STC, low-dimensional chaos, and turbulence, tools such as power spectrum [82], filtering spatiotemporal diagrams [88], embedding dimension, and time series analysis [89, 90] are inadequate. A classification of these phenomena has been reported in the literature [91–97].

Lyapunov spectrum

The only reliable tool for the characterization of STC is the LS which is composed of a set of LE. In the case of STC, the LS has a continuous set of positive values. This matches the definition that has been proposed in [91, 93]. In the case of low dimensional chaos, the LS possesses a discrete set of positive values. However, the turbulence or weak turbulence is characterized by a power law cascade of a scalar quantity [98].

The computation of the LS itself is very well documented [99] and is not the purpose here. Let us just recall the main steps. From the state of the system at a given time, the linear evolution of any small perturbation δX can be described by $\partial_t \delta X = J\delta X$, where J is the respective Jacobian. In the present case, we introduce $\psi = \psi_r + i\psi_i$, with ψ_r and ψ_i being the real and imaginary part of ψ respectively. At a time $t = t_0$, introducing $\psi = \psi_0 + \delta\psi$, with $\delta\psi \ll \psi(t = t_0) = \psi_0$, the matrix J reads:

$$J = \begin{bmatrix} -(1 + 2\psi_{0r}\psi_{0i}) & \Delta - \psi_{0r}^2 - 3\psi_{0i}^2 - \partial_\tau^2 \\ -\Delta + (\psi_{0i}^2 + 3\psi_{0r}^2) + \partial_\tau^2 & -(1 - 2\psi_{0r}\psi_{0i}) \end{bmatrix}, \tag{1.24}$$

and $\delta X = (\delta\psi_r, \delta\psi_i)^t$. Suppose that we want to compute the n^{th} first dominant exponents of the spectrum, we introduce the matrix L, which contains n orthonormal vectors v_i which to be used as initial conditions when solving $\partial_t \delta X = J\delta X$:

$$L(t = t_0) \equiv [v_1 \quad v_2 \quad \ldots \quad v_n] = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{d1} & x_{d2} & x_{d3} & \cdots & x_{dn} \end{bmatrix}, \tag{1.25}$$

where d is the dimension of the system. After a time increment dt, the matrix L evolves to $L(t_0 + dt) = \hat{U}L(t_0)$ where $\hat{U} = e^{J\cdot dt}$. Using the modified Gram-Schmidt QR decomposition on $L(t_0 + dt)$, the diagonal elements of R account for the LE $\tilde{\lambda}_i (i = 1, \ldots, n)$ at time $t_0 + dt$, that is:

$$\tilde{\lambda}_i(t_0 + dt) = \frac{1}{dt} \ln(R_{ii}(t_0 + dt)). \tag{1.26}$$

After repeating this procedure several times, after a large number of iterations N, the LE can be approximated by:

$$\lambda_i \equiv \langle \tilde{\lambda}_i \rangle = \frac{1}{N dt} \sum_{k=1}^{N} \ln(R_{ii}(t_0 + k dt)). \tag{1.27}$$

Kaplan-Yorke dimension

The Kaplan-Yorke Dimension (D_{KY}) is a crucial tool for characterizing the complexity of chaotic attractors in spatiotemporal systems [100]. It serves as an order parameter to establish the bifurcation diagram of STC. The D_{KY} is defined from the spectrum of LE $\{\lambda_i\}$ as:

$$D_{KY} = p + \frac{\sum_{i=1}^{p} \lambda_i}{|\lambda_{p+1}|} \tag{1.28}$$

where p is the largest integer that satisfies $\sum_{i=1}^{p} \lambda_i > 0$ [32]. The D_{KY} grows linearly with the volume of a high-dimensional chaotic system [36, 92], meaning that for a one-dimensional system of size L, STC implies that D_{KY} increases linearly with L. In the context of the Lugiato-Lefever Equation (LLE), we consider a temporal window, ΔT, and for a 1D system:

$$D_{KY} = \xi_{\delta}^{-1} \Delta T \tag{1.29}$$

where ΔT represents the temporal extension of the system and ξ_{δ} represents the Lyapunov dimension density of the system for a fixed value of the control parameter. This intensive quantity provides an estimation of the extension of dynamically independent subsystems generated by the chaotic dynamics. For a fixed set of parameters, it is useful to provide an intensive characterization of the chaoticity level by computing the slope of the D_{KY} curve with respect to the volume [101, 102]. The inverse of this slope, also known as the Lyapunov dimension density (ξ_{δ}), estimates the size of the independent sub-domains produced by the presence of the attractor. Studies have shown that the range of independent sub-domains decreases with the pumping level [103]. This also demonstrates the evolution of the D_{KY} as a function of the temporal window for several pump powers and showed that the curves' slopes increase with the pump power, confirming the spatiotemporal chaotic nature of the process. In Ref. [101], they also showed the dependence of this D_{KY} on the control parameter, which confirms the increasing of the complexity with the pump power. Besides this growth of the D_{KY} with the pump power, we must also consider the conjecture that the dimension increases with the size of the system–extensive nature of the STC [32, 91, 92, 95, 104].

1.4 Extreme Events

Our research centers on a synchronously pumped fiber ring cavity, a type of dissipative system known for exhibiting Extreme Events (EE) induced by Spatiotemporal Chaos [101]. Notably, this study [101] marks the first documentation of EE arising within the monostable regime of a passive Kerr-type cavity. EE are very rare and large amplitude fluctuations that arise not just by chance but through the complex interplay of forces within systems far from equilibrium. In the field of optics, these events manifest as rare and intense optical pulses and have analogies in hydrodynamics, notably with oceanic rogue waves, which are both described by the Nonlinear Schrödinger Equation (NLSE) [105], illustrating a common theoretical foundation. Most optical studies have occurred in fibers where the interplay of nonlinearity, dispersion, and noise can generate EE [2–4, 106].

From a statistical perspective, EE lie in the tails of probability distributions, which describe the likelihood of occurrences of varying magnitude. These tails can be 'heavy' or 'sub-exponential', which means that the probability of EE does not decrease as rapidly as in a standard Gaussian distribution. This statistical perspective allows scientists to estimate the probability of such events or the probability of an event exceeding a specific threshold.

Dynamically, EE are not simply the result of chance but result from the complex interplay of forces within complex systems far from equilibrium. This complexity is evident in various environments, from weather systems modeled on fundamental physical principles to the chaotic dynamics observed in financial markets and traffic systems.

In recent years, EE generated by complex systems have been a topic of great interest in various fields due to their significant impact on society [107], on scientific research, and on our daily lives. The presence of EE extends beyond optics and oceanography [108] and can also be observed in other disciplines, such as astrophysics [109], the atmosphere [110], and geology [111].

Although significant progress has been made, predicting these events remains challenging due to their complex dynamics and the high dimensionality of the systems involved [107]. Regardless of their different physical origins, these EE can be viewed as emergent phenomena, where information coming from one physical system can benefit others. Optics, in particular, offers many advantages in this respect, allowing well-controlled laboratory conditions, good detection techniques, and fast timescales for large data collection samples in a limited time, facilitating the study of nonlinear dynamics and the development of predictive models. The ability to predict such events holds great practical importance, enhancing the performance and reliability of systems in weather forecasting, traffic management, power grids, and financial markets. A variety of methods have been employed to predict EE in the time series of chaotic DS. Some of the approaches that have proven effective include nonlinear dynamics estimation based on the Koopman operator theory [112] and Takens embedding theorem [113], to Machine Learning (ML) approaches such as support vector machines [114], singular spectrum analysis, the maximum entropy method [115], and the advanced deep learning techniques, including auto-encoders [116], Long Short-Term Memory (LSTM) networks [11], and reservoir computing [117] all contributing significantly to developing models capable of predicting chaotic systems with high-dimensional attractors. In the following sections, we will introduce model-free tools like Machine Learning (ML) and Information Theory tools to help predict these extreme events.

1.5 Machine Learning

Machine Learning (ML), a subset of Artificial Intelligence (AI), uses statistical techniques and numerical algorithms to perform tasks without explicitly programmed instructions. Examples of ML tasks include classification, detection, pattern recognition, prediction, optimization, and modeling complex dynamics from observed data. These capabilities make ML valuable across various domains, including control systems, speech processing, neuroscience, and computer vision. Recently, its application has expanded to predicting chaotic systems and optimizing optical systems. ML algorithms are broadly categorized into two types: supervised and unsupervised learning.

- Supervised learning is still the most common approach in current ML [118]. It uses prior knowledge to build models that describe system responses based on input-output relationships. In other words, it refers to an approach where a model maps output values Y from the inputs X. The training of the model is realized by introducing pairs of input and corresponding known output values (X_n, Y_n). The predictions of the model, denoted as $\hat{Y}$, are compared to the ground truth Y, and the parameters of the model are modified in a way that the prediction error is minimized based on the pairs of input and output values.

- Unsupervised learning involves exploring data to identify inherent patterns without predefined labels.

Both learning paradigms are instrumental in analyzing complex datasets and extracting meaningful insights. In our work, we use only supervised learning methods.

Recently, there has been a growing interest in applying ML techniques to optical systems. The application of ML in optics and photonics is relatively recent, but it has shown significant promise in various areas like laser optimization [6, 7], ultrashort pulse measurements [8], label-free cell classification [119], imaging [120–122] and coherent communications [123]. ML techniques have been used to predict EE in optical systems, such as rogue solitons, by correlating spectral measurements with temporal peak intensities. This review [5] provides an overview of the application of ML in photonics, its different applications, and the challenges faced in this interdisciplinary field.

1.5.1 Neural Networks

Neural Network (NN) have become the workhorse in most Artificial Intelligence (AI) applications over the last decade. Early applications of Neural Network (NN) for modeling and predicting DS date back to the work of Lapedes and Farber [124], who demonstrated the effectiveness of feedforward Artificial Neural Networks (ANN) in modeling deterministic chaos. Krischer et al. used

Artificial Neural Networks (ANN) to forecast the principal components of a spatiotemporal catalytic reaction [125]. As an alternative to ANN, wavelet networks were proposed by Cao et al. for chaotic time series prediction [126]. However, these early approaches were often limited to low-dimensional systems and typically employed in conjunction with dimensionality reduction techniques.

Recurrent Neural Networks (RNN) have the potential to overcome these scalability issues and be applied to high-dimensional spatiotemporal dynamics. RNN are a special type of NN, commonly employed in processing sequential or time-dependent data, such as time series [127], speech recognition [128, 129], and language translation [130–132]. Unlike traditional feedforward NN and classical numerical methods that aim at discretizing existing equations of complex systems, RNN models are data-driven and have recurrent connections that allow them to maintain a memory of previous inputs through hidden states propagated over time. This is key for identifying patterns and learning dynamics that evolve sequentially over time [133]. The architecture of an RNN is illustrated in Fig. 1.8, it includes nodes with recurrent connections that store information from previous input values. The hidden state at any given time step is a function of the current input and the hidden state from the previous time step.

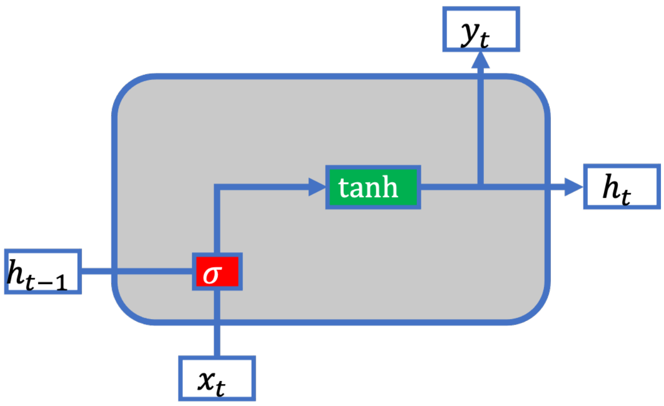


Figure 1.8: Schematic representation of a RNN. The RNN processes sequential data by maintaining a hidden state that captures information from previous time steps. This hidden state h_t is updated at each time step based on the current input x_t and the previous hidden state h_{t-1}.

The works of Takens [134] and Sauer [21] demonstrated that the dynamics on a D-dimensional attractor of a dynamical system could be unfolded in a time-delayed embedding of dimension greater than 2D. The identification of a useful embedding and the construction of a forecasting model has been a significant area of research [135]. More recently, a data-driven method using the Koopman operator formalism [136] was proposed [137] with feed-forward ANN to identify an embedding space with linear dynamics suitable for theoretical analysis.

There is limited work at the interface of RNN and non-linear dynamical systems [9–11, 138, 139]. RNN, designed to capture long-term dependencies in sequential data [140–143], initially faced challenges such as vanishing or exploding gradients. The development of LSTM networks, which use gates to manage memory, significantly improved RNN performance by addressing these issues [127] and have been one of the standard RNN approaches for the past 15 years. In recent years, RNN architectures have been benchmarked for various applications, such as short-term load forecasting in supply networks [144] and extreme event detection in low-dimensional time series [145]. LSTM networks have also been used as surrogates to model kinematics in fluid flows [146] and to capture long-term statistics in reduced-order spaces of dynamical systems [11]. Furthermore, they have been applied to model residual dynamics in Galerkin-based reduced-order models [139] and to forecast chaotic chimera states [147]. A more recent development is the GRU model [148], introduced in 2014, which simplifies the LSTM architecture. In the following sections, we will introduce both models and provide a brief comparison between them.

1.5.2 Long Short-Term Memory networks

Long Short-Term Memory (LSTM) introduce memory cells and gating mechanisms that regulate the flow of information, allowing the network to maintain and update its hidden state more effectively. This architecture enables LSTM to learn long-term dependencies and perform well in tasks requiring the retention of information over extended periods. The operation of an LSTM cell [127] is shown in Fig. 1.9. Panel (a) demonstrates the overall operation of the cell, where the output y_t at a given time t of a data sequence is calculated based on the cell input x_t (at time t), the hidden state h_{t-1} and the cell state c_{t-1} from the previous time step. The output of the cell therefore takes into account long-term information about the previous inputs and outputs of the cell that are internally stored in the cell state. Panel (b) shows a more detailed description of the cell operation. The equations describing the operations of a single LSTM cell at time t for an input $x_t \in \mathbb{R}^{d_i}$ can be written as:

$$\begin{aligned} f_t &= \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) & \tilde{c}_t &= \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) \\ i_t &= \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) & c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ o_t &= \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) & h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{1.30}$$

Here, vectors f_t, i_t, and $o_t \in \mathbb{R}^{d_h}$ are the forget, input, and output gates, respectively, with d_h denoting the dimensionality of the hidden state, i.e., the number of hidden units. Vector $\tilde{c}_t \in \mathbb{R}^{d_h}$ is the cell input activation, and vectors c_t and $h_t \in \mathbb{R}^{d_h}$ are the updated cell and hidden states, respectively. Matrices $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_o$, and $\mathbf{W}_c \in \mathbb{R}^{d_h \times (d_h + d_i)}$ are the cell weights, which are the parameters that the model learns during training. They are used to transform the input and the previous hidden state before applying the activation functions. Biases $\mathbf{b}_f$, $\mathbf{b}_i$, $\mathbf{b}_o$, and $\mathbf{b}_c \in \mathbb{R}^{d_h}$ are essential for shifting the activation function's output, allowing the model to fit the training data more effectively. The symbol $\odot$ denotes pointwise Hadamard multiplication, and σ is the sigmoid function, which is an activation function that squashes the input to a value between 0 and 1. This function is crucial for the gates in an LSTM cell because it allows the network to make decisions about which information to keep and which to discard. By outputting a value close to 0 or 1, the sigmoid function effectively acts as a switch, either letting information pass through or blocking it. The mathematical expression for the sigmoid function is: $\sigma(x) = \frac{1}{1+e^{-x}}$. Notation $[\mathbf{h}_{t-1}, \mathbf{x}_t]$ indicates the concatenation of the two vectors. The weights and biases of the network are iteratively trained via backpropagation through time [149].

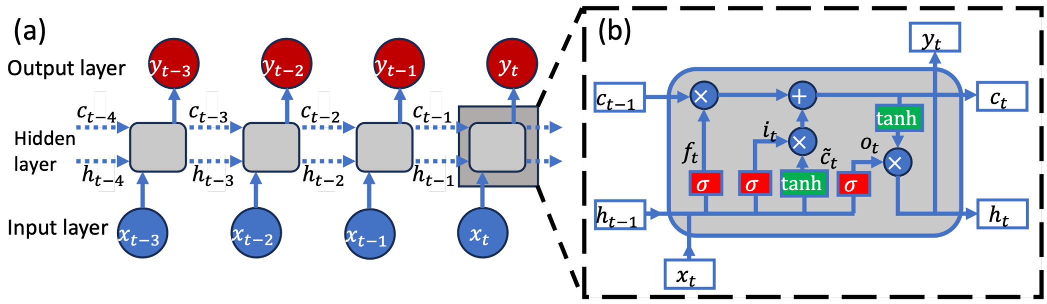


Figure 1.9: Illustration of LSTM cell operation. (a) Represents an overview of LSTM cell operation in the context of a sequence, showing how the output y_t at time t is influenced by the input x_t, hidden state h_{t-1}, and cell state c_{t-1}. (b) Detailed schematic of the internal operations within an LSTM cell, highlighting the forget gate f_t, input gate i_t, cell input activation $\tilde{c}_t$, cell state c_t, output gate o_t, hidden state h_t, and output gate y_t.

1.5.3 Gated Recurrent Units

Building on the LSTM architecture, Gated Recurrent Unit (GRU) [148] combine the forget gate and input gate into a single update gate (z_t), and merge the hidden state with cell state, making the architecture simpler, reduce the computational complexity and improve the training speed. There are only two gates in the cell architecture of GRU: the reset gate (r_t), which decides how much past information to forget, and the update gate (z_t), which works in a similar way to the forget and input gates of an LSTM. It defines how much of the previous memory remains, which can be described by Eqs. 1.31 and 1.32, respectively:

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \tag{1.31}$$

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \tag{1.32}$$

where W_r is the weight matrix of the reset gate, b_r is the bias matrix of the reset gate, W_z is the weight matrix of the update gate, and b_z is the bias matrix of the update gate. Subsequently, the new memory cell state is obtained by Eq. 1.33:

$$\tilde{h}_t = \tanh(W_h[h_{t-1} \odot r_t, x_t] + b_h) \tag{1.33}$$

where W_h is the weight matrix of the new memory cell state and b_h is the bias matrix of the new memory cell state. The gating signal (z_t) ranges from 0 to 1. The closer the gating signal is to 1, the more data will be memorized, whereas the closer to 0, the more data will be forgotten. Therefore, one single expression can control both forgetting and inputting, generating the output (h_t):

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{1.34}$$

Fig. 1.10 illustrates the GRU cell architecture, showing the reset gate (r_t) and the update gate (z_t), as well as the process for generating the new memory cell state ($\tilde{h}_t$) and the final output (h_t).

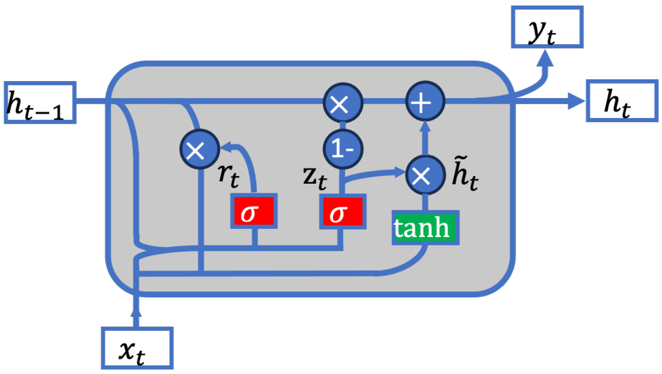


Figure 1.10: Illustration of a GRU cell. The figure shows the reset gate (r_t) and the update gate (z_t), as well as the process for generating the new memory cell state ($\tilde{h}_t$) and the final output (h_t).

1.6 Information Theory and Transfer Entropy

Information theory, established by Claude Shannon [150], provides a mathematical framework for quantifying information transfer, uncertainty, and dependency between random variables. This field has profound implications in various domains, including communications, cryptography, data compression, and complex system analysis. It allows us to understand and measure how information is transmitted and transformed within systems.

1.6.1 Fundamental Measures of Information Theory

Key measures in information theory include Shannon entropy, joint entropy, conditional entropy, mutual information, and Conditional Mutual Information (CMI) [151]. These measures help quantify the uncertainty and relationships between different variables in a system.

- **Shannon Entropy** measures the uncertainty or randomness of a random variable X [150, 152]. It is defined as:

$$H(X) = -\sum_{x \in X} P(x) \log P(x) \tag{1.35}$$

  For example, consider a binary set $X = \{0, 1, 0, 1, 1, 0, 0, 1\}$ with equal probabilities for 0 and 1. By calculating the probabilities: $p(0) = \frac{4}{8} = 0.5, \quad p(1) = \frac{4}{8} = 0.5$.
  The entropy $H(X)$ is: $H(X) = -(0.5 \log_2 0.5 + 0.5 \log_2 0.5) = 1$ bit, would be maximal, indicating high uncertainty.

- **Joint Entropy** $H(X, Y)$ extends the concept of entropy to two variables, measuring the uncertainty in their combined states:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} P(x, y) \log P(x, y) \tag{1.36}$$

  If we have another binary set $Y = \{1, 0, 1, 0, 0, 1, 1, 0\}$, the joint entropy $H(X, Y)$ captures the combined uncertainty of X and Y. The joint probabilities might be: $p(0,0) = 0.25, \quad p(0,1) = 0.25, \quad p(1,0) = 0.25, \quad p(1,1) = 0.25$. The joint entropy $H(X, Y)$ is: $H(X, Y) = -(0.25 \log_2 0.25 + 0.25 \log_2 0.25 + 0.25 \log_2 0.25 + 0.25 \log_2 0.25) = 2$ bit.

- **Conditional Entropy** $H(X|Y)$ measures the amount of uncertainty remaining in one variable given that the value of another variable is known.

$$H(X|Y) = -\sum_{x \in X} \sum_{y \in Y} P(x, y) \log P(x|y) \tag{1.37}$$

where $P(x|y)$ is the conditional probability that $X = x$ given $Y = y$. If knowing Y reduces the uncertainty in X, the conditional entropy $H(X|Y)$ would be lower. For our binary sets, the conditional entropy $H(X|Y)$ is: $H(X|Y) = -(0.25\log_2 1 + 0.25\log_2 1 + 0.25\log_2 1 + 0.25\log_2 1) = 0$ bit. This means that knowing Y completely determines X, so there is no uncertainty remaining.

- **Mutual Information** $I(X;Y)$ between X and Y measures the average reduction in uncertainty about x that results from learning the value of y, or vice versa [152, 153], as follows:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \tag{1.38}$$

It highlights the dependency and shared information between X and Y. It can also be expressed in terms of entropy:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{1.39}$$

Given the sequences X and Y and their respective probabilities:

$$I(X;Y) = H(X) - H(X|Y) = 1 - 0 = 1 \text{ bit.}$$

This means that knowing Y reduces the uncertainty about X by 1 bit, indicating a strong dependence between X and Y. If $I(X;Y) = 0 \Leftrightarrow X$ is independent of Y.

- **Conditional Mutual Information (CMI)** $I(X;Y|Z)$ measures the amount of information that one random variable contains about another, given that a third variable is known [154]. For binary variables, it is defined as:

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) \tag{1.40}$$

where $H(X|Y,Z)$ is the conditional entropy of X given Y and Z. In other words, it quantifies how much knowing Y reduces the uncertainty about X, given that Z is known.

1.6.2 Transfer Entropy

Transfer Entropy is a non-parametric statistic introduced by Thomas Schreiber in 2000 [155], which builds on the fundamental measures of information theory shown in the previous section to quantify the directed transfer of information between two processes. It is particularly useful for identifying causal relationships in complex systems, capturing both linear and non-linear dependencies.

The TE from a source process Y to a target process X is defined as:

$$T_{Y\to X} = \sum P(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log \frac{P(x_{t+1}|x_t^{(k)}, y_t^{(l)})}{P(x_{t+1}|x_t^{(k)})} \tag{1.41}$$

where t is a time index, $x_t^{(k)}$ and $y_t^{(l)}$ represent the k and l past values of x and y, up to and including time n (with $k, l = 1$ being default choices). This formulation captures how the future state of X can be predicted from the past states of Y, indicating the directional influence Y has on X.

To enhance the accuracy of TE, it is important to consider the possibility that the influence of Y on X might occur with a source-target lag or delay 'τ' time steps [156]. The delayed TE is expressed as:

$$T_{Y\to X}(\tau) = \sum P(x_{t+1}, x_t^{(k)}, y_{t-\tau}^{(l)}) \log \frac{P(x_{t+1}|x_t^{(k)}, y_{t-\tau}^{(l)})}{P(x_{t+1}|x_t^{(k)})} \tag{1.42}$$

where τ represents the delay between the influence of Y on X. This addition allows for a more accurate representation of the dynamics in systems where interactions are not instantaneous.

When the history length is set to 1 ($k = 1, l = 1$), TE simplifies to CMI [157]. Specifically, if X and Y are single-step time series, TE can be expressed as:

$$T_{Y\to X} = I(X_{t+1}; Y_t|X_t) \tag{1.43}$$

Here, TE is equivalent to the CMI between the future state of X and the past state of Y, given the past state of X. However, when the history length is greater than 1 ($k > 1$ or $l > 1$), TE accounts for more complex, multi-step dependencies and thus extends beyond CMI, capturing richer dynamics in the interactions between X and Y.

TE will be the model-free tool used in multiple chapters of this thesis. In practice, there are many codes that allow the computation of the TE of continuous time series. We chose the open-source Java Information Dynamics Toolkit (JIDT) software package by Lizier [158] (`https://github.com/jlizier/jidt/`). The portability of this JIDT Java-based code, with no installation requirement, has motivated our choice, which we employed in Matlab. The TE is computed between either two univariate or multivariate time series of observations using Kraskov-Stögbauer-Grassberger (KSG) estimation. The calculation is performed by examining K^{th} nearest neighbors in the joint distribution [159, 160] rather than two mutual information calculators as initially suggested by Kraskov [161]. The value given by the TE is in 'nats' (with 1 nat being equal to $1/\ln(2)$ bits).

Before TE, Norbert Wiener proposed that a time series Y causes X if the inclusion of past values of Y improves the prediction of X. Clive Granger formalized this in 1969 into what is known as Granger causality, based on the premise that if Y Granger-causes X, then past values of Y should contain information that helps predict X beyond the information contained in past values of X alone [162]. Unlike Granger causality, which is parametric and typically assumes linear relationships, TE is non-parametric and model-free, capable of capturing nonlinear interactions and providing a more flexible and accurate representation of information dynamics in complex systems. TE provides a robust framework for studying information dynamics and causal relationships in both linear and nonlinear systems, offering advantages over traditional methods like Wiener and Granger causality. Its ability to capture delayed interactions further enhances its utility in understanding complex temporal dependencies.

In addition, TE is especially powerful in studying spatiotemporal causality, involving the directional transfer of information across both space and time in a system. This is crucial in many fields, such as meteorology, neuroscience, and ecology, where the dynamics of a system at one location or time point can significantly influence other locations or time points. For example, in meteorological studies, TE can be used to analyze how wind speed and direction at one meteorological station can affect conditions at another station. By examining the TE between these stations, we can identify patterns of information flow and causal influences across different spatial locations. This helps in understanding and predicting weather patterns by capturing the complex, nonlinear interactions that occur in the atmosphere. The ability of TE to capture nonlinear dependencies makes it particularly suited for studying these complex systems. Traditional methods like Granger causality may fail to detect these intricate relationships because they often assume linear interactions. In contrast, TE does not make such assumptions, allowing it to reveal the true dynamics of information transfer. Overall, TE provides a comprehensive tool for investigating spatiotemporal causality and the transportation of nonlinear information between different regions, making it invaluable for studying and predicting the behavior of complex systems.

Chapter 2

Spatiotemporal Chaos Forecasting in an Optical Fiber Ring Resonator

Outline of the current chapter

In this chapter, we explore the predictability of spatiotemporal chaos dynamics. Our investigation centers on numerical data exhibiting spatiotemporal chaos, analyzed through various Machine Learning (ML) techniques. The predictive methodologies are divided into two primary approaches. Initially, we focus on forecasting the complete spatiotemporal dynamics. This involves employing advanced ML models such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) to predict the system's future states based on current and historical data. However, our results indicate that long-term forecasting in these chaotic systems is inherently difficult due to their sensitive dependence on initial conditions. Given these challenges, we propose an alternative method termed "local forecasting." This approach focuses on predicting specific future events by identifying precursors. Using statistical analysis and information theory, we detect signals or patterns that precede significant events within the chaotic data. These precursors enhance the precision of our event-specific forecasts, allowing for a more targeted and potentially more reliable prediction method. Through these approaches, we aim to advance the understanding and forecasting capabilities of complex spatiotemporal systems.

2.1 Introduction

Many recent advances in the understanding of complex dynamics have been driven by experimental and theoretical studies in modern optics, in fields as diverse as optical-fiber cavities with quadratic or cubic nonlinear materials subject to externally injected radiations [163–165]. The field of optics, therefore, is ideally suited to the investigation of Spatiotemporal Chaos (STC) and turbulence in dissipative systems far from thermodynamic equilibrium (cavity and laser systems). Within a few decades, optical Kerr-nonlinear resonators have emerged as the paradigmatic setup for the study of externally driven nonlinear systems [63, 166, 167]. Kerr resonators are also known for the property to continuously switch between monostable (a region where the transmission function is single-valued for a given pump power) and bistable (S-shape transmission curve) regimes that we talked about in Section 1.3.1. Operating out of equilibrium, Kerr resonators can exhibit nontrivial outputs such as cavity solitons [168–170] and Modulation Instability (MI) [68, 171] that we talked about in Sections 1.3.4 and 1.3.3 respectively.

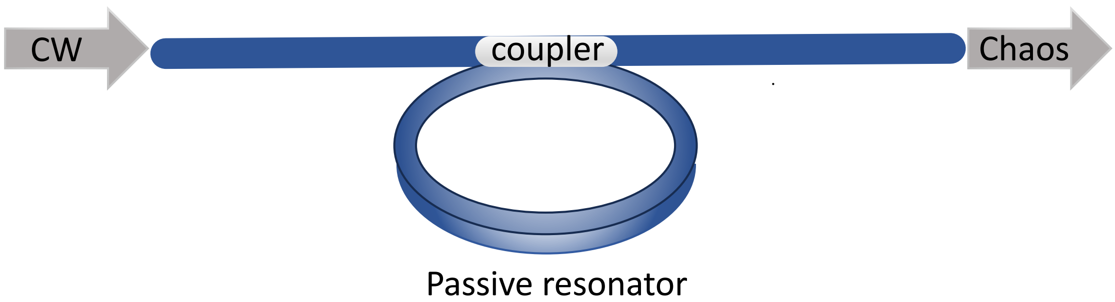


Figure 2.1: Schematic representation of the optical fiber ring. The Continuous-Wave (CW) input is coupled into the passive resonator, leading to a chaotic output.

In our study, we consider a passive resonator made of an optical fiber ring synchronously pumped close to a cavity resonance. The ring was set to operate in a monostable regime. By pumping the cavity well above the cavity threshold, typically a few times, the continuous wave solution breaks into a periodic wave train, which in turn experiences an oscillatory instability and then evolves into a chaotic regime [101, 172], as shown in Fig. 2.1. This current sequence is universal and can be observed in many other fields of physics [32, 173]. The dynamic of the light circulating in the cavity is accurately modeled by the driven and damped Nonlinear Schrödinger Equation (NLSE) augmented with boundary conditions, in the mean field limit, referred to as Lugiato-Lefever Equation (LLE) [58] which we already showed in Section 1.3:

$$\frac{\partial \psi(t,\tau)}{\partial t} = S - (1 + i\Delta)\psi - i\eta \frac{\partial^2 \psi}{\partial \tau^2} + i \mid \psi \mid^2 \psi \tag{1.12}$$

Here $\psi(t,\tau)$ represents the normalized slowly varying envelope of the electric field circulating within the cavity, S denotes the pump strength, Δ is the frequency detuning, t corresponds to the slow evolution of ψ over successive round-trips, and τ accounts for the fast dynamics that describe how the electric field envelope changes along the fiber. The cavity is also set to operate in the anomalous dispersion regime so $\eta = -1$. The homogeneous ($\frac{\partial^2 \psi_S}{\partial \tau^2} = 0$) and steady state ($\frac{\partial \psi_S}{\partial t} = 0$) solutions of Eq. 1.12 obey the cubic equation:

$$|S|^2 = I_s[1 + (I_s - \Delta)^2] \tag{1.17}$$

Here $I_s = \mid \psi_S \mid^2$ is the intracavity field intensity.

2.2 Numerical Simulation

We perform a numerical simulation of the LLE as presented in Eq. 1.12 and illustrated in Fig. 2.2. The simulation employs a numerical integration approach that combines the Fourth Order Runge-Kutta (RK4) method with a three-point Finite Difference Method (FDM). The parameters chosen for the simulation are as follows: Detuning $\Delta = 1.1$, which positions the system in a monostable regime since $\Delta < \sqrt{3}$, power losses $\alpha = 0.2$, for a pump $S = 3.3$ and intensity $I_s = 2.8$. The initial condition is set as the steady state field, $\psi_s = \sqrt{I_s} e^{i\phi_s}$ (Eq. 1.15), with $\phi_s = \tan^{-1}(I_s - \Delta)$ (Eq. 1.16), and is perturbed by a small fluctuation to initiate dynamics in the system. We define the scaled time variable as $t = \frac{\alpha t'}{t_R}$, where t' represents the simulation time (Eq. 1.11c). The time step $\delta t = 0.01$ ensures that

20 time steps occur per cavity round trip. The simulation covers a total duration from $t = 0$ to $t = 2250$, corresponding to 11250 cavity round trips. While this number of round trips is sufficient for achieving statistically meaningful results in our study, it represents a very small duration when considered in an experimental context. The temporal window for the simulation spans from $\tau_r = -64$ to $\tau_r = 64$ with $N = 1024$ discrete points, resulting in a temporal resolution of $\delta\tau = \frac{2\times\tau_r}{N} = 0.125$. This temporal window captures the dynamics over each round trip within the cavity.

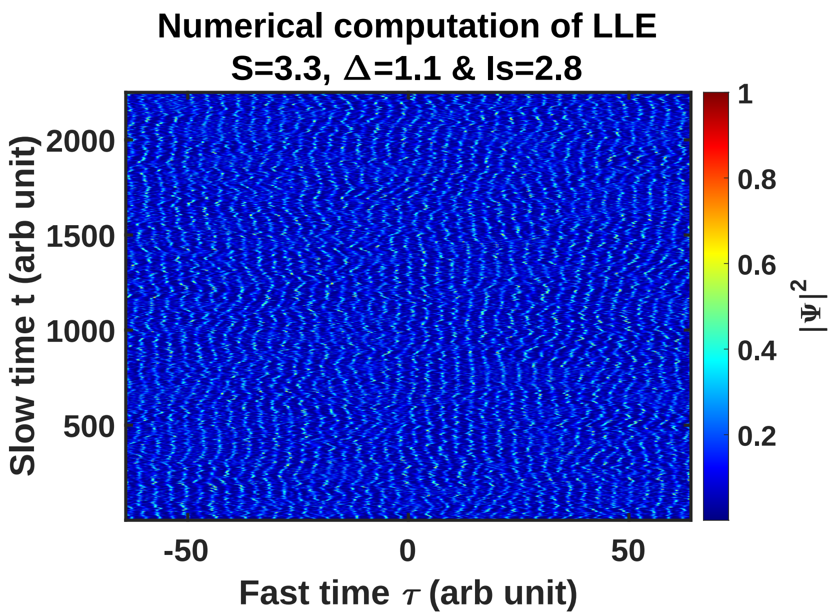


Figure 2.2: Numerical computation of LLE showing the evolution of the field intensity $|\psi|^2$ along the fiber over time. Parameters: $S = 3.3$, $\Delta = 1.1$, $I_s = 2.8$.

To better visualize the evolution of the field intensity over time, Fig. 2.3 provides a zoomed-in view of the same data presented in Fig. 2.2. This zoomed-in view allows for a clearer observation of the spatiotemporal structures and dynamics. An almost periodic pulse train can be observed. Pulse positions and shapes modifications in this two-dimensional map are characteristic of a STC [101].

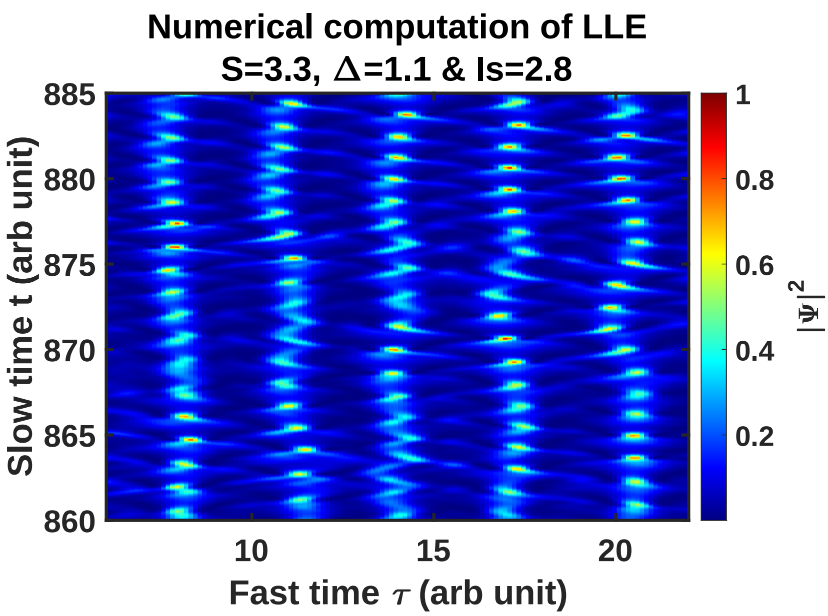


Figure 2.3: Zoomed-in view of figure 2.2.

2.3 Identification and Characterization of Spatiotemporal Chaos

In this section, we revisit the work done by Coulibaly et al. [103] on the identification and characterization of the STC observed in this system, using several key metrics such as the Lyapunov Spectrum (LS) and the Kaplan-Yorke Dimension (D_{KY}). We already discussed in detail these two metrics in the previous chapter (Section 1.3.6), and in this study [103] they showed that this dynamic has a LS with a positive continuous part which is indicative of STC. For this high-dimensional chaotic system, the D_{KY} grows linearly with the system's volume, confirming the extensive nature of the chaos.

We refer to Figure 3 from Coulibaly et al.'s work [103] (Fig. 2.4 here), which provides a comprehensive characterization of the STC in Kerr resonators. The figure shows experimentally and numerically the STC observed in the Kerr resonator, highlighting the irregular patterns in the output field. The LS demonstrates a positive continuous part, a plot showing the D_{KY} as a function of the

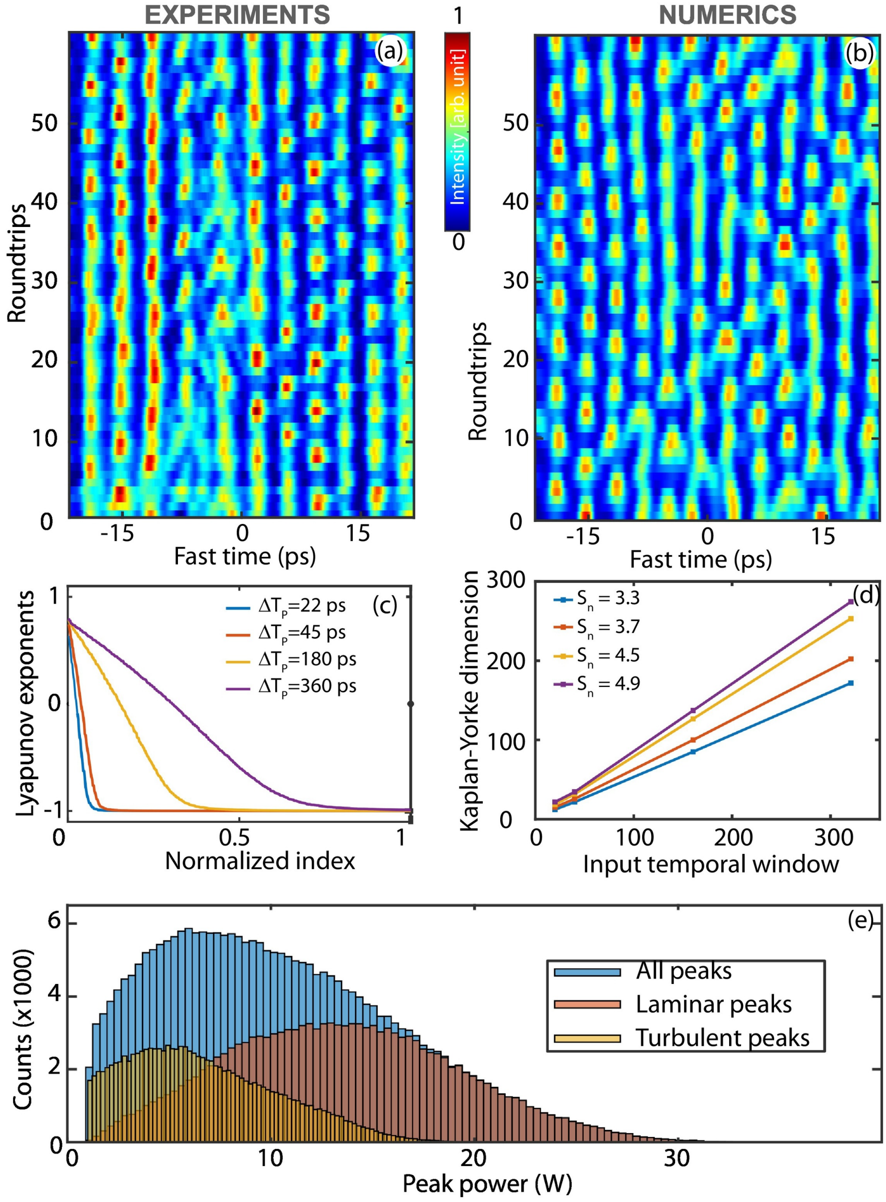


Figure 2.4: Characterization of STC: (a) Experimental results showing STC, (b) Numerical simulations, (c) LS, (d) D_{KY}, and (e) Probability density functions of the peaks. Fig. taken from [103].

temporal window, illustrating the linear growth with the system's volume.

2.4 Analogy with Hydrodynamics

As we said earlier in Section 1.4, the description of the STC can be achieved by analogy with hydrodynamics [101]. The dynamic shown in Fig. 2.3 is an irregular succession of laminar and turbulent flows. A detailed statistical study of the laminar or turbulent domains was performed in [101] and it was done following the process described in [174]. The probability distribution of the laminar/turbulent domains has the following mixture function: $P(x) = (Ax^{-\mu} + B)e^{-mx}$. All the constants depend only on the parameters except m which also changes with the value of the power set to separate the laminar and turbulent domains. The bursts detected during the evolution can be labeled according to their location in a laminar or turbulent flow, respectively. Distributions of all the bursts, those located in laminar and turbulent domains are shown in Fig. 2.4(e) for a threshold set at the mean value of the intracavity power. An important observation here is that the highest bursts are mainly located in the laminar flows. In Fig. 2.5, we show a turbulent and a laminar region from our map and we can see that the highest bursts are located in the laminar region. In the turbulent region (Fig. 2.5a), we can see an irregular and uneven pattern with low power pulses, while in the laminar region (Fig. 2.5b) it consists of a well-defined spatiotemporal pattern consisting of a train of pulses with high power that are identified as EE.

2.5 Forecasting The Full Dynamic

The forecasting of high-dimensional chaotic systems has taken a big step forward after the improvements in supervised ML algorithms. These studies have been done mainly using Deep Learning (DL) and RNN. The ability of these NN to find hidden correlations in the data has allowed us to make predictions about complex spatiotemporal dynamics. By providing model-free processes, it is possible that the tools of chaos theory are no longer needed to deal with time series in general. In this study, our objective is to forecast the full spatiotemporal chaotic dynamic of a high-dimensional non-linear system generated by the passive resonator using two types of RNN: the LSTM and the GRU discussed

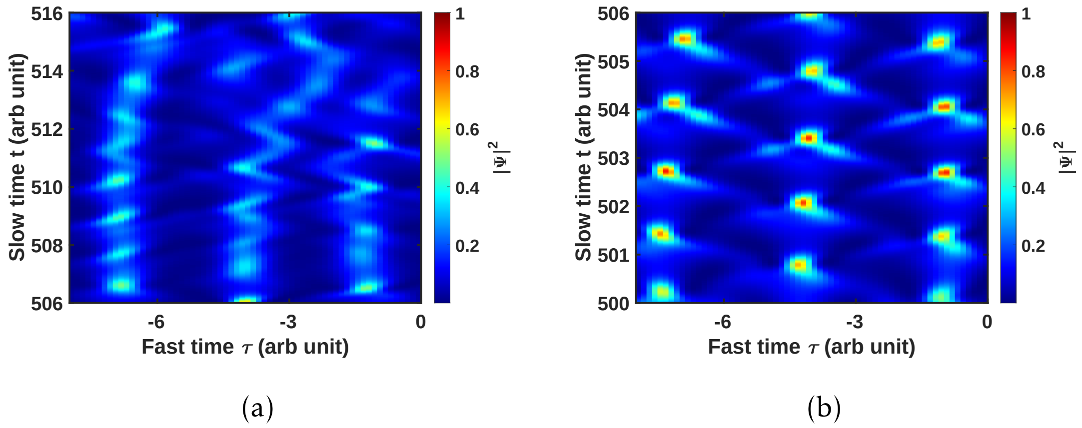


Figure 2.5: Visualization of turbulent and laminar regions. (a) Turbulent region showing an irregular and uneven pattern with low power pulses. (b) Laminar region displaying a well-defined spatiotemporal pattern with a train of high power pulses.

in Sections 1.5.2 and 1.5.3, respectively, which were found to be well-suited for sequence-to-sequence predictions due to their ability to maintain state information over extended sequences, which is crucial for capturing the underlying dynamics of chaotic systems.

Considering the increase in complexity with system size in high-dimensional chaotic systems, which makes the dynamic difficult to predict, we begin our analysis with a scaled-down model. Specifically, we reduce the system size from $N = 1024$ to $N = 128$ and adjust τ_r to 8 to maintain equivalent temporal resolution. Instead of simulating the system over a total duration from $t = 0$ to $t = 2250$, corresponding to 11250 cavity round trips, we limit our simulation to $t = 225$, which corresponds to 1125 round trips. This modified system setup is illustrated in Fig. 2.6, providing a simpler yet sufficiently complex framework to test our forecasting methodologies effectively.

2.5.1 Data preparation for Full Dynamic Forecasting

To forecast the full spatiotemporal chaotic dynamic, it is essential to properly prepare the data for training, testing, and forecasting using the LSTM and GRU networks. This involves selecting appropriate segments of the data and

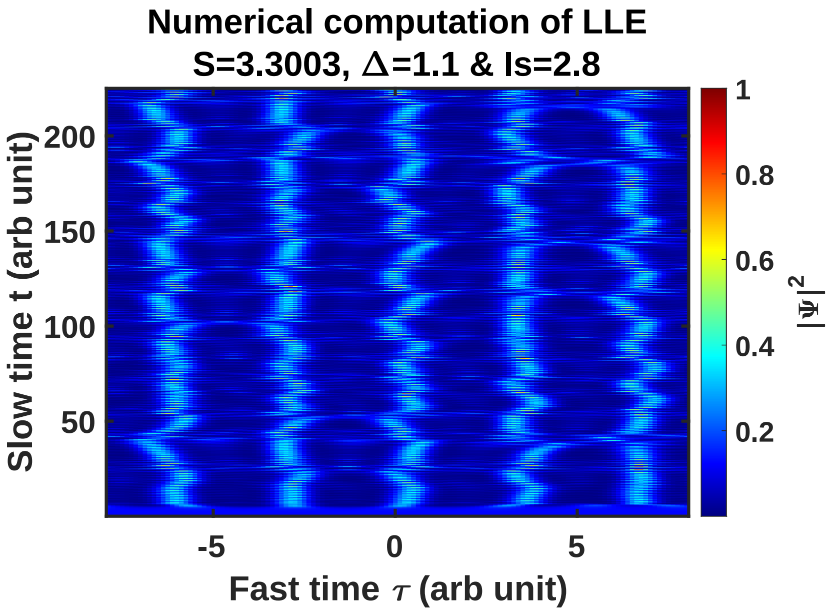


Figure 2.6: Numerical computation of LLE showing the evolution of the field intensity $|\psi|^2$ along the fiber over time for smaller system size (size $N = 128$). Parameters: $S = 3.3$, $\Delta = 1.1$, $I_s = 2.8$.

structuring them in a format suitable for these networks. The data is extracted from numerical simulations of the LLE, as illustrated in Fig. 2.7, where the training, testing, and forecasting datasets are highlighted by red, yellow, and green squares, respectively. The datasets are referred to as `x_train`, `x_test`, and `x_forecast`. Below are the detailed steps involved in preparing these datasets:

1. **Exclude the Transitory Region**: The initial transitory phase (indicated below the red square in the figure) is removed to ensure that the networks train on stable, representative data.

2. **Select `x_train`**: Start `x_train` immediately after the transitory region and include data up to 70% of the remaining dynamic length excluding the length equivalent to a lag, a 'lag' refers to a round-trip delay. `x_train` is reshaped to `(x_train.shape[0], 1, x_train.shape[1])` to match the network input format, and the rest of the datasets are reshaped in a similar way. This ensures that each sequence fed into the network represents the spatial evolution of

the field over one time step.

3. **Select `y_train`**: `y_train` is identical to `x_train` but shifted upward by the lag duration. This means `y_train` will start after the beginning of `x_train` by some time steps equal to the lag and end at the same number of time steps after `x_train` ends.

4. **Select `x_test`**: `x_test` starts immediately after the end of `x_train` and spans the next 10% of the dynamic length.

5. **Select `y_test`**: `y_test` is similar to `x_test` but with respect to `y_train`.

6. **Select `x_forecast`**: `x_forecast` starts right after the end of `x_test` and includes the final 20% of the dynamic length, adjusted for the lag.

7. **Select `y_forecast`**: `y_forecast` starts right after the end of `y_test` and corresponds to `x_forecast` shifted by the lag.

This structured approach ensures that each sequence fed into the network represents the spatial evolution of the field over one time step. Our goal is to train the neural network using `x_train` and `y_train`, test its performance on `x_test` and `y_test`, and use it to forecast future states from `x_forecast`, and compare the predicted outcomes with `y_forecast`.

2.5.2 Models Architecture for Full Dynamic Forecasting

The codes were written in Python using Keras library [175] with TensorFlow backend [176], representing the state of the art for deep neural network training and prediction. Fig. 2.8 shows the architecture of the LSTM and GRU models used for forecasting the full spatiotemporal chaotic dynamics of the system. Both the LSTM and GRU models share several common architectural components designed for effective sequence-to-sequence prediction. Each model consists of the following elements:

1. **Model Initialization**: Both models are initialized as sequential models, allowing layers to be added one after another in a linear stack.

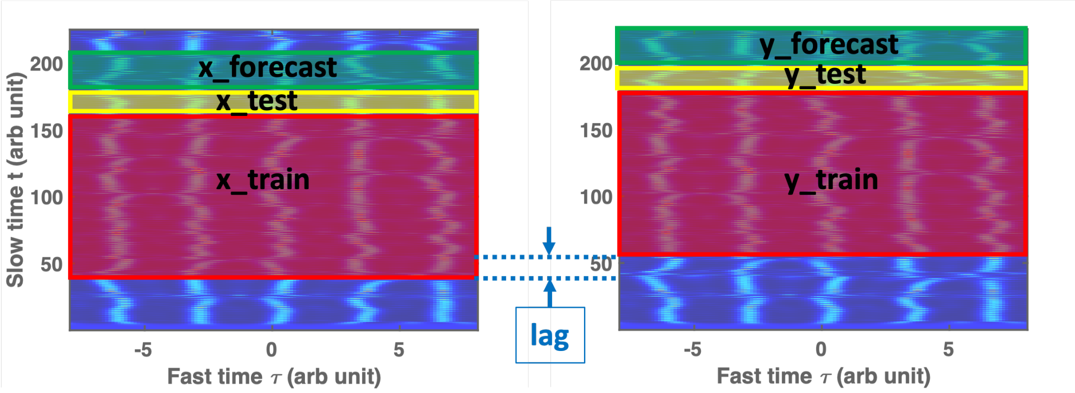


Figure 2.7: Illustration of data selection for training, testing, and forecasting sets. The left panel shows the input data: the training set in red (`x_train`), the testing set in yellow (`x_test`), and the forecasting set in green (`x_forecast`). The right panel shows the corresponding output data shifted by a lag of many time steps: `y_train`, `y_test`, and `y_forecast`, respectively. This lag is illustrated by the blue dashed arrows. The difference between the input and output panels indicates the forecasted time steps.

2. **Recurrent Layers**: Each model uses recurrent layers (either LSTM or GRU) with hidden units and the Rectified Linear Unit (ReLU) activation function. Hidden units in these layers play a crucial role in capturing the temporal dependencies and dynamics of the input sequences. The number of hidden units, `nhidden`, determines the capacity of the model to learn from the data. More hidden units can capture more complex patterns, but they also increase the computational cost and the risk of overfitting. The first recurrent layer in each model is configured to return sequences, ensuring that the output includes all hidden states for each time step.

3. **Repeat Vector Layer**: This layer repeats the input from the previous layer times to match the desired output sequence length, which is necessary for the decoder part of the sequence-to-sequence model.

4. **Dropout Layers**: Dropout layers are included after the recurrent layers to prevent overfitting by randomly setting a fraction of input units to 0 during training. This regularization technique helps improve the generalization capability of the model by preventing it from becoming too specialized for

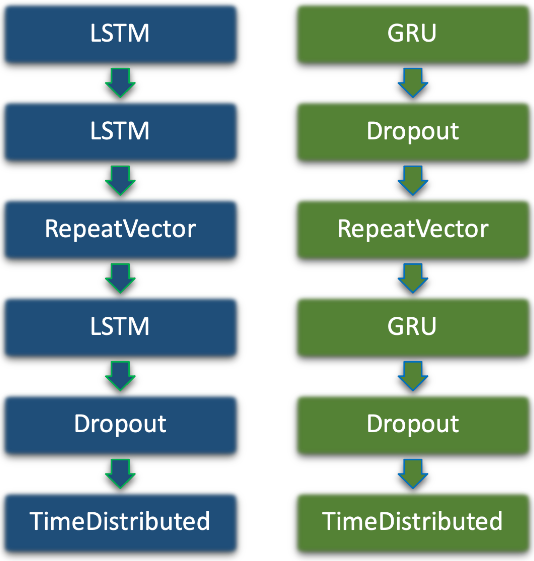


Figure 2.8: Architecture of the LSTM and GRU models used for forecasting the full spatiotemporal chaotic dynamics of the system. The LSTM model (left) consists of three LSTM layers, followed by a dropout layer, and a final TimeDistributed layer. The GRU model (right) consists of two GRU layers, each followed by a dropout layer, and a final TimeDistributed layer. Both models include a RepeatVector layer to handle sequence-to-sequence prediction.

the training data.

5. **Time Distributed Dense Layer**: This layer applies a fully connected Dense layer to each time step individually, ensuring that each time step of the output sequence is processed independently. The number of features in the output converts the output of the recurrent layers into the final desired output format, which can be compared to the true values during training.

The primary differences between the models are the type and number of recurrent layers.

2.5.3 Training and Evaluation of the Networks for Full Dynamic Forecasting

After defining the architecture of the LSTM and GRU models, the next step is to compile and fit the models to the training data.

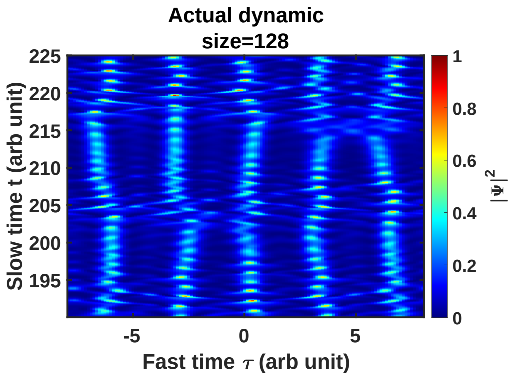
For the compilation step, we used the Huber loss, which is less sensitive to outliers in data than the mean squared error (MSE) loss, combining the best properties of mean absolute error (MAE) and MSE. This makes it well-suited for our data, which includes a small fraction of very high values. As an optimizer, we chose the Adam optimizer for its efficiency and adaptability [177]. We evaluated model performance using the RMSE during training. The RMSE measures the square root of the average squared differences between predicted and actual values, providing a robust indication of prediction accuracy.

For fitting the model, the training was performed using `x_train` as input sequences and `y_train` as target sequences. The model was trained for 100 epochs, with an epoch representing one complete pass through the entire training dataset. A batch size was used to determine the number of samples propagated through the network at each step, balancing computational efficiency and model performance. Callback functions were employed during training to adjust the learning rate and implement early stopping. Specifically, a learning rate scheduler was included to dynamically adjust the learning rate, decreasing it by 25% every 10 epochs. This reduction helps the model to fine-tune its parameters more delicately as training progresses. Early in training, a higher learning rate allows the model to make significant updates to the weights, which speeds up convergence. As training continues, a lower learning rate helps to stabilize the learning process, ensuring that the model does not overshoot the optimal parameter values and can settle into a more precise minimum of the loss function. Early Stopping was employed to prevent overfitting, which stops training if the validation loss does not improve for several consecutive epochs. The training data was shuffled before each epoch to enhance the model's robustness and prevent it from learning the order of the training data. Validation data (`x_test` and `y_test`) was used to evaluate the model's performance after each epoch, helping to monitor the model's generalization ability and detect overfitting. The training and validation loss curves are plotted to visualize the model's learning progress. These plots, which will be shown in Section 2.5.5, provide insights into how well the model fits the training data and its generalization to unseen data.

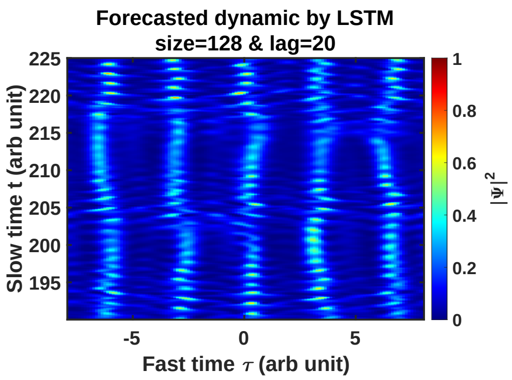
2.5.4 Forecasting the Full Dynamic Using the LSTM and GRU Models

Once the LSTM and GRU models have been trained and validated, the next step is to employ these models to forecast the spatiotemporal chaotic dynamics of the system. The prediction step involves using the trained models to forecast future values from the input data `x_forecast`. The resulting forecasts by the LSTM and GRU models, `y_forecast_lstm` and `y_forecast_gru`, respectively, allow us to evaluate the accuracy of the models by comparing them with the actual future values (`y_forecast`). This comparison helps in understanding how well the models perform in forecasting the complex dynamics of the system. To demonstrate the effectiveness of these models across different conditions, we analyze their performance for various system sizes ($N = 64, 128, 256$) and temporal lags. In the following, we focus on demonstrating the impact of round-trip delays on forecasting accuracy by comparing the outcomes at two specific lags: 20, corresponding to a single round-trip delay, and 200, equivalent to 10 round-trip delays, for a system size of 128.
Fig. 2.9 illustrates the actual dynamic alongside the forecasted dynamics by the LSTM and GRU models for a system size of 128 and for both lag periods. From the analysis of the figure, the models generally predict the dynamics accurately for a lag of 20. However, for a lag of 200, while the overall patterns and trends of the chaotic system are captured, there are noticeable discrepancies, particularly in predicting higher values. These discrepancies indicate a potential reduction in prediction accuracy with an increase in round-trip delay, suggesting that the model's ability to handle long-term dependencies may be limited.

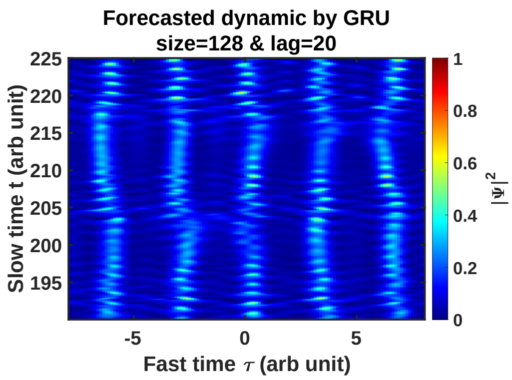
In the upcoming section, we will evaluate the performance of the LSTM and GRU models in more detail, focusing on differences in accuracy between the LSTM and GRU models across various metrics and conditions. This will provide a more comprehensive understanding of each model's strengths and weaknesses in forecasting spatiotemporal chaotic dynamics.



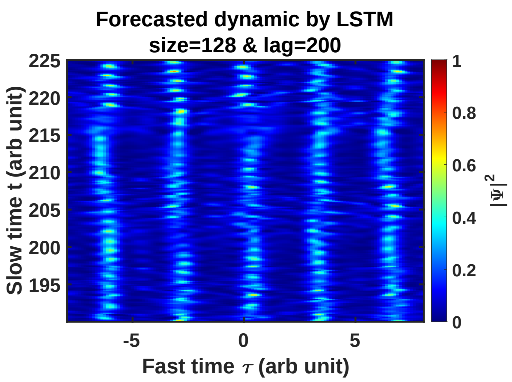
(a)



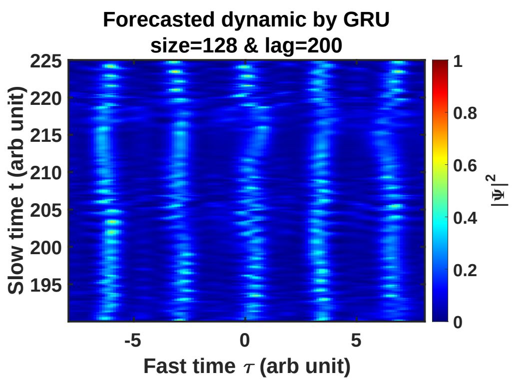
(b)



(c)



(d)



(e)

Figure 2.9: (a) Actual dynamics, forecasted dynamics by (b) LSTM and (c) GRU for a lag of 20 (one round-trip delay), by (d) LSTM and (e) GRU for a lag of 200 (10 round-trips delay) for the high-dimensional chaotic system with $I_s = 2.8$.

2.5.5 Evaluating Model Accuracy for Full Dynamic Forecasting

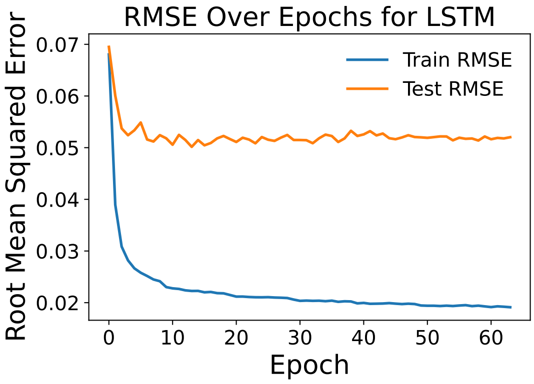
To evaluate the accuracy of our LSTM and GRU models, we used several metrics, including Root mean squared error (RMSE) loss over epochs and Pearson correlation (PC). These metrics provide a comprehensive understanding of how well our models predict the spatiotemporal chaotic dynamics.

Root mean squared error Loss Over Epochs

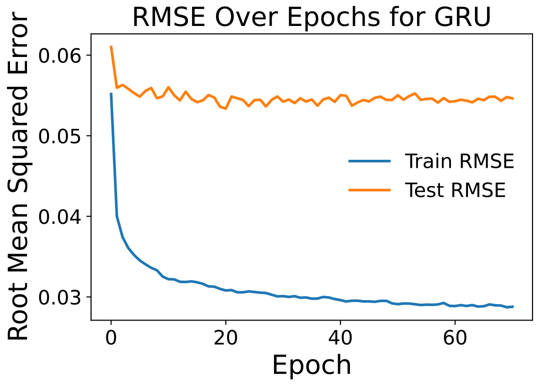
The Root mean squared error (RMSE) is a crucial metric for evaluating the accuracy of our LSTM and GRU models in predicting the spatiotemporal chaotic dynamics. It measures the differences between the predicted and actual values during training and validation, providing a robust indication of prediction accuracy. Lower RMSE values indicate better model performance. To assess the model performance, we plotted in Fig. 2.10 the RMSE loss for both training and validation sets over epochs for different lags (20 and 200). These plots help us understand how well the models are learning the underlying patterns and generalizing to unseen data. For lag 20, both LSTM and GRU models show a rapid decrease in RMSE for the training set, indicating that the models quickly learn the underlying patterns. The validation RMSE also decreases significantly and stabilizes, showing that the models generalize well to unseen data. This indicates good model performance with the ability to accurately predict future values based on the input data. However, for lag 200, the scenario changes. While the training RMSE continues to decrease, indicating that the models are fitting the training data well, the validation RMSE remains higher and exhibits more fluctuation. This suggests that the models may be overfitting the training data and are unable to generalize effectively to new data. The higher RMSE and instability in the validation loss over epochs reflect reduced model performance for longer forecasting horizons.

Pearson correlation

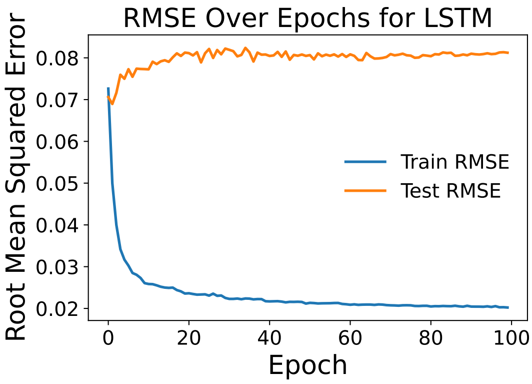
The Pearson correlation (PC) metric provides insights into how well our models predict the spatiotemporal chaotic dynamics. It measures the linear



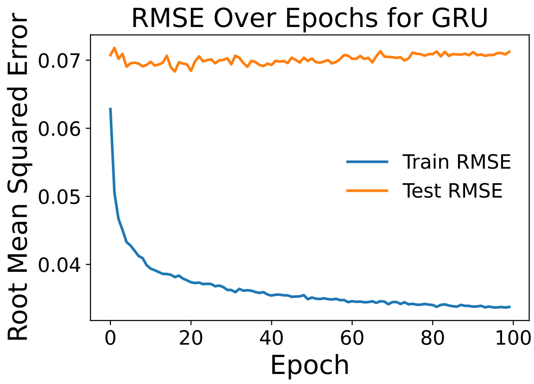
(a)



(b)



(c)



(d)

Figure 2.10: RMSE Loss Over Epochs for LSTM and GRU for system size 128 and $I_s = 2.8$. (a) and (b) Lag 20: Lower RMSE indicates good model performance, with convergence observed early in training. (c) and (d) Lag 200: Higher RMSE indicates reduced model performance, with less stability in the validation loss over epochs.

relationship between the predicted and actual values, providing insights into how well the models capture the trends and patterns in the data. A PC coefficient close to 1 indicates a strong positive linear relationship, while a coefficient close to 0 indicates no linear relationship. In our analysis, we evaluated the performance of both models using the Maximum (MAX) and Standard deviation (STD) of the predicted and actual values. By computing the PC coefficient for both the MAX and STD, we can determine how well the forecasted values align with the actual measurements.

- For lag 20, both LSTM and GRU models exhibit high PC coefficients (Figs. 2.11a and 2.11b respectively), indicating a strong linear relationship between forecasted and actual values. This high correlation suggests that the models effectively capture the underlying dynamics and can predict future values with high accuracy. Specifically, the LSTM model shows a PC of 0.93 for the STD and 0.92 for the MAX, while the GRU model shows a PC of 0.86 for both the STD and MAX.

- However, for lag 200, the PC coefficients are significantly lower, indicating a weaker linear relationship and reduced predictive accuracy. This decrease in PC reflects the challenges associated with long-term forecasting in chaotic systems, where the models struggle to maintain accuracy over extended prediction horizons. The LSTM model exhibits a PC of 0.41 for the STD and 0.45 for the MAX, while the GRU model shows a PC of 0.45 for both the STD and MAX (Figs. 2.11c and 2.11d respectively).

In Fig. 2.11, we can visualize the plot of the forecasted values against the actual values for both models and for different lags. The plots include a black line representing perfect predictions ($y = x$), which helps in assessing how closely the forecasted values match the actual values. By using these evaluation metrics, we can comprehensively evaluate the performance of our LSTM and GRU models, ensuring they accurately capture the dynamics of the high-dimensional chaotic system. This detailed analysis of the Pearson correlation helps us understand the strengths and limitations of our models, guiding further improvements and optimizations for better forecasting performance.

2.5.6 Changes in Pearson Correlation for Different System Sizes and round-trip delays

As shown earlier, PC is a crucial metric for assessing the accuracy of our LSTM and GRU models in forecasting spatiotemporal chaotic dynamics. To comprehensively evaluate the performance of the LSTM and GRU models, we analyzed PC across different system sizes, round-trip delays, and system complexities. Specifically, we considered complexities $I_s = 2.2$, representing a less

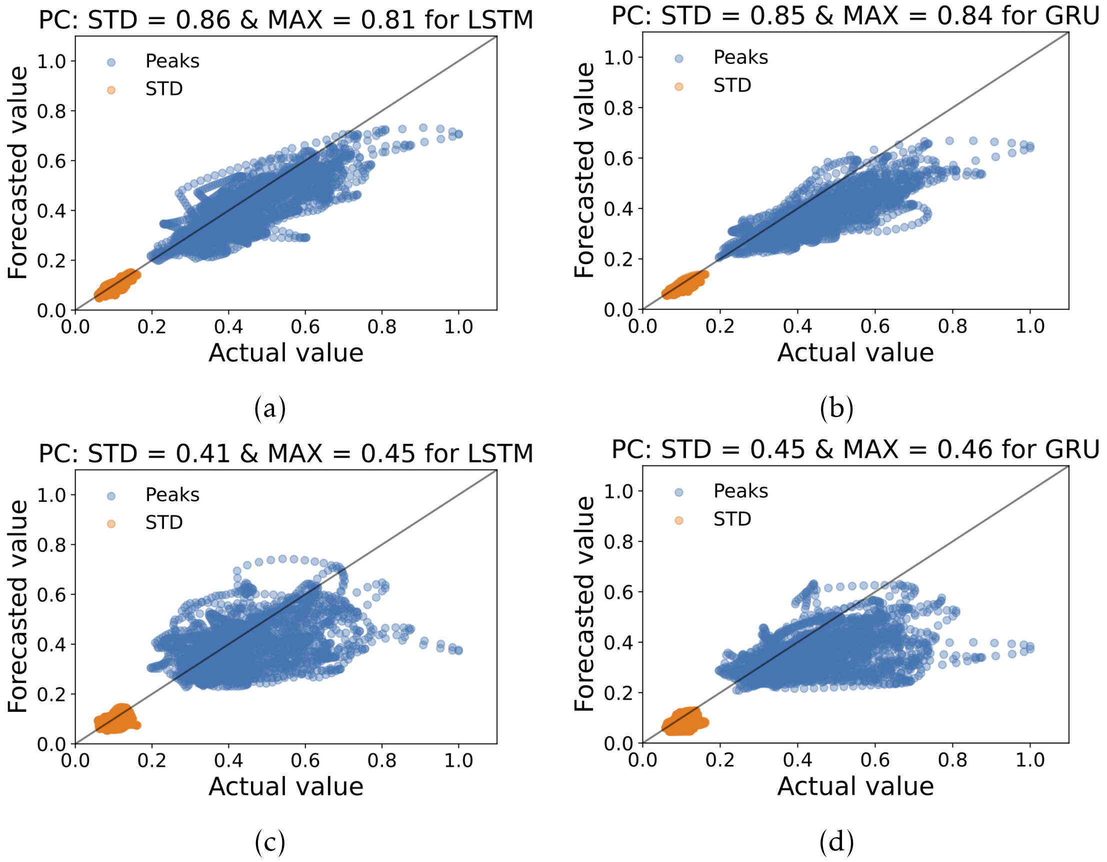


Figure 2.11: PC for LSTM and GRU for system size 128 and $I_s = 2.8$. (a) and (b) Lag 20: High PC indicates strong correlation. (c) and (d) Lag 200: Lower PC indicates weaker correlation. The scatter plots show the relationship between the forecasted and measured values, with the black line representing a perfect prediction ($y = x$). The blue dots represent the peaks, and the orange dots represent the standard deviations.

complex system, and $I_s = 2.8$, indicative of a more complex system. Fig. 2.12 shows the PC for the LSTM and GRU models as a function of lag for system sizes 64, 128, and 256.

- For a less complex system ($I_s = 2.2$), as illustrated in Fig. 2.12a, the PC remains consistently high across all round-trip delays, indicating that both LSTM and GRU models maintain strong linear relationships between forecasted and actual values. This suggests that the models can effectively capture the dynamics of less complex systems, even as round-trip delays increase.

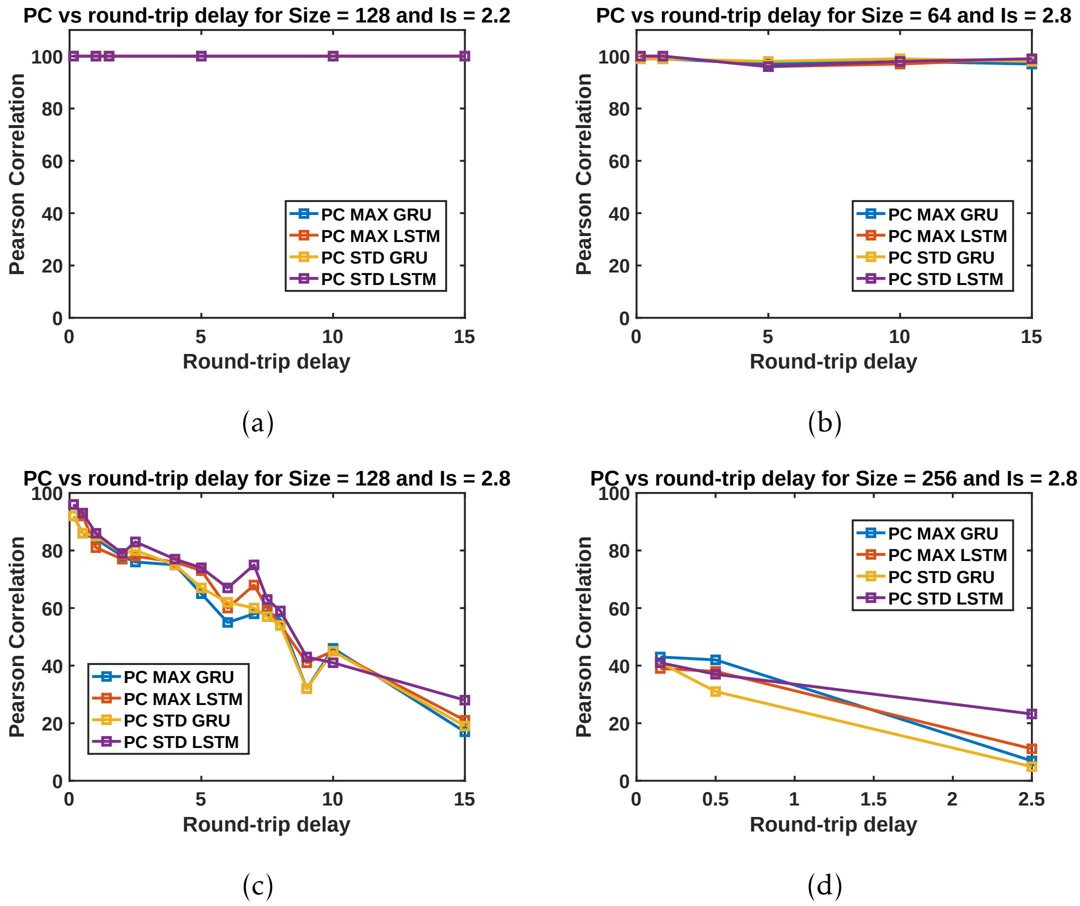


Figure 2.12: PC for LSTM and GRU models across different system sizes and complexities. (b) Size 64 and $I_s = 2.8$: PC remains high across all lags, indicating robust prediction accuracy due to the smaller system size. (a) Size 128 and $I_s = 2.2$: PC remains high across all lags for a less complex system. (c) Size 128 and $I_s = 2.8$: PC decreases as the lag increases, indicating reduced accuracy for more complex systems. (d) Size 256 and $I_s = 2.8$: PC decreases more significantly with increasing lag, reflecting the challenge of modeling larger and more complex systems.

- The high accuracy observed for system size 64 ($I_s = 2.8$), as shown in Fig. 2.12b, suggests that smaller systems are inherently easier for the models to forecast accurately across all lags.

- However, as we increase the system size to 128, we notice a marked decrease in PC with increasing lag values, as depicted in Fig. 2.12c. An important observation is that beyond a round-trip delay of 9, the PC falls below 50% for the two models. This significant drop suggests that the accuracy of forecasting future dynamics deteriorates sharply, marking a critical limit beyond which the current models struggle to provide reliable predictions.

- For the largest system size of 256 ($I_s = 2.8$), detailed in Fig. 2.12d, the PC decreases even more significantly as the lag increases, reflecting the substantial challenges involved in accurately modeling larger and more complex systems.

These observations underscore the importance of considering system size, lag, and complexity when evaluating the performance of LSTM and GRU models. By understanding how these factors impact PC, we can better assess the models' ability to generalize and accurately forecast the spatiotemporal chaotic dynamics in various scenarios. The marked decline in forecasting accuracy for long-term predictions of the full spatiotemporal chaos dynamic in larger and more complex systems, such as those developed by the fiber ring cavity, underscores the challenge posed by high dimensionality and highlights the need for further model improvements or alternative forecasting approaches.

2.6 Local Forecasting of Extreme Events in Spatiotemporal Chaos

An alternative method to forecasting the full spatiotemporal chaos dynamic is "local forecasting". This method addresses the crucial questions of when and where extreme events will emerge within chaotic regimes, as well as answering the question of what is coming by forecasting the profile and location of the upcoming event based on previously identified precursors. In the context of spatiotemporal chaos, a precursor is defined as a detectable signal or pattern that occurs prior to an extreme event, providing an early indication of its imminent occurrence. These precursors are identified by analyzing specific patterns in the spatiotemporal data that precede significant events. They are captured at defined time lags across multiple cavity round trips. Recent research has extensively explored the use of precursors for predicting extreme events in chaotic systems. For instance, Coulibaly et al. demonstrated the effectiveness of using machine learning to predict chaotic extreme pulses in Kerr resonators by identifying precursors from spatiotemporal data [103]. Similarly, Pammi et al. highlighted the predictive capabilities of nonlocal partial information in a spatiotemporally chaotic microcavity laser, showing that precursors were crucial for accurate forecasting of extreme events [117]. These studies underscore the significance of precursors in forecasting extreme events and provide a foundation for the methods employed in this work. Additionally, the use of precursors has been supported by other researchers across different chaotic systems, further validating its utility [178, 179]. We chose this approach because extreme pulses are observed in the laminar region, where we have an almost periodic pulse train, which facilitates the identification of the precursors (as shown in Fig. 2.5b).

The method combines model-based tools such as the two-point correlation function with model-free tools from information theory and Deep Neural Network (DNN). This combination simplifies the overall forecasting task and enhances the accuracy of the predictions.

The initial step in this process is the identification of precursor-pulse pairs using the spatial and temporal two-point correlation function [36, 180, 181]. To optimize the determination of the size of subdomains necessary for this identification and to justify the choice of precursors over bursts, we compute the information flow using the 2D map of the TE (which we introduced in Section 1.6.2). We begin by characterizing the spatiotemporal dynamics through these two-point correlation functions and verifying the information flow with the TE. Next, we detect all precursors within the dynamic and associate each precursor with its corresponding pulse to create a set of precursor-pulse pairs. This dataset is then split into training, test, and validation sets. The training set is used to train sequence-to-sequence forecasting RNN models, which are designed to use the sequential data of identified precursor-pulse pairs to predict future pulses.

2.6.1 Slow and Fast Time Periodicity

In dynamic systems exhibiting complex evolutions, understanding the coherence between different spatial and temporal locations is crucial. This coherence is quantitatively assessed through the equal-time two-point correlation function, both spatially and temporally. These functions, defined as follows, measure the probability that two points separated by $\delta\tau$ (space) and δt (time) will behave coherently:

$$C(\delta\tau) = \langle(\psi(\delta\tau + \tau', t) - \langle\psi\rangle)(\psi(\tau', t) - \langle\psi\rangle)\rangle. \tag{2.1}$$

$$C(\delta t) = \langle(\psi(\tau, \delta t + t') - \langle\psi\rangle)(\psi(\tau, t') - \langle\psi\rangle)\rangle. \tag{2.2}$$

The brackets $\langle\cdot\rangle$ stand for the average process. However, the direct determination of $C(\delta\tau)$ and $C(\delta t)$ is computationally intensive. To address this, we apply the Wiener-Khinchin theorem [182, 183], which simplifies the process by computing the time-average of the Fourier spectra, followed by an inverse Fourier transform of its magnitude squared. The two functions are plotted in Figs. 2.13 and 2.14.

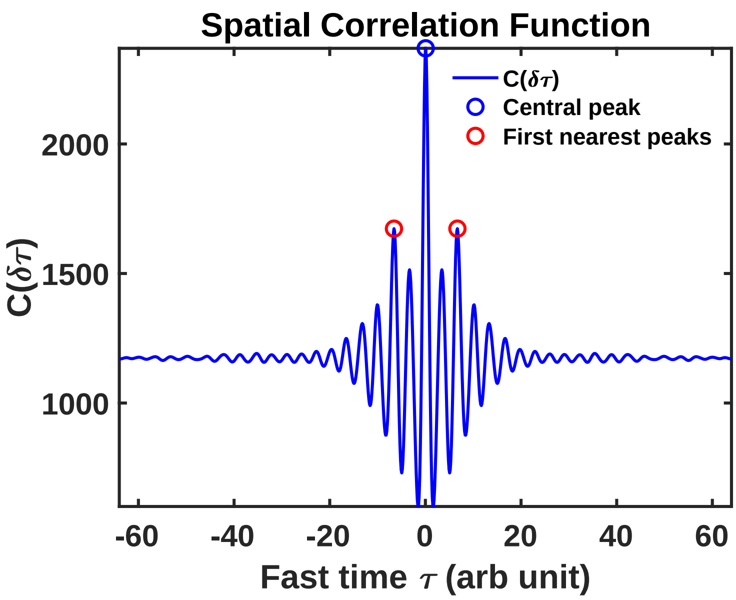


Figure 2.13: Spatial correlation function illustrating the central peak and adjacent peaks. The adjacent peaks indicate a repeating pattern in space.

In Fig. 2.13, the spatial correlation function reveals a central peak at zero separation, signifying maximum correlation. The adjacent peaks indicate a repeating pattern in space, confirmed by observations in laminar regions (see Fig. 2.5b). The separation to the first significant secondary peak measures the Fast time periodicity (FTP), equivalent to 6.7 (arb unit) or 52 discrete points.

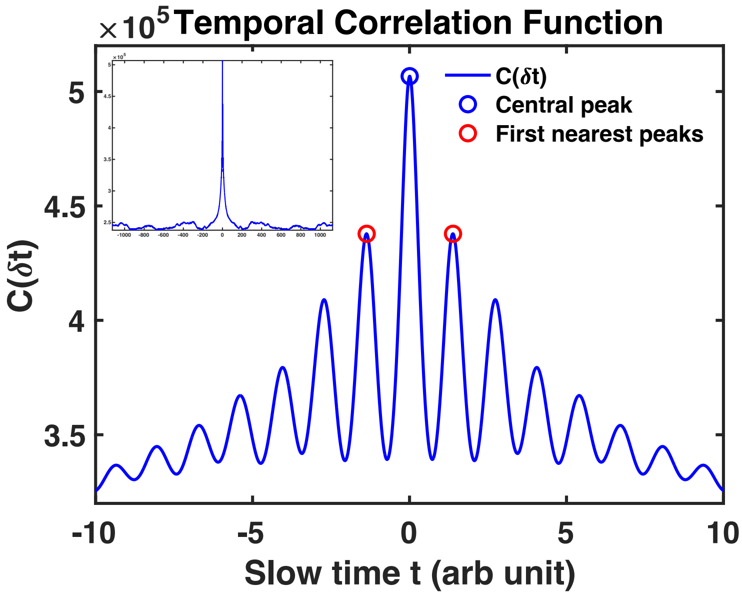


Figure 2.14: Zoomed-in view of the inset figure, representing the temporal correlation function showing central and secondary peaks, indicative of periodic bursts.

Similarly, Fig. 2.14 displays the temporal correlation function, where the periodicity between significant events is captured. The distance between the central peak and the nearest secondary peaks defines the Slow time periodicity (STP) equivalent to 1.37 simulation time units or 6.8 cavity round trips.

2.6.2 Transfer Entropy 2D Map

To optimize the determination of the size of subdomains necessary for the identification of the precursors, the information flow by the 2D map of the TE is computed.

We compute $\mathrm{TE}_{Y\to X}(\delta t, \delta\tau)$ with $X \equiv |\psi(t,\tau)|^2$ and $Y \equiv |\psi(t-\delta t, \tau+\delta\tau)|^2$, ψ being the considered field, as sketched in Fig. 2.15. With the TE given by:

$$\mathrm{TE}_{Y\to X} = \sum_{x,y} p(x_{n+1}, x_n^h, y_n^h) \log\left[\frac{p(x_{n+1}|x_n^h, y_n^h)}{p(x_{n+1}|x_n^h)}\right] \tag{2.3}$$

with n the current iteration and h the history length.

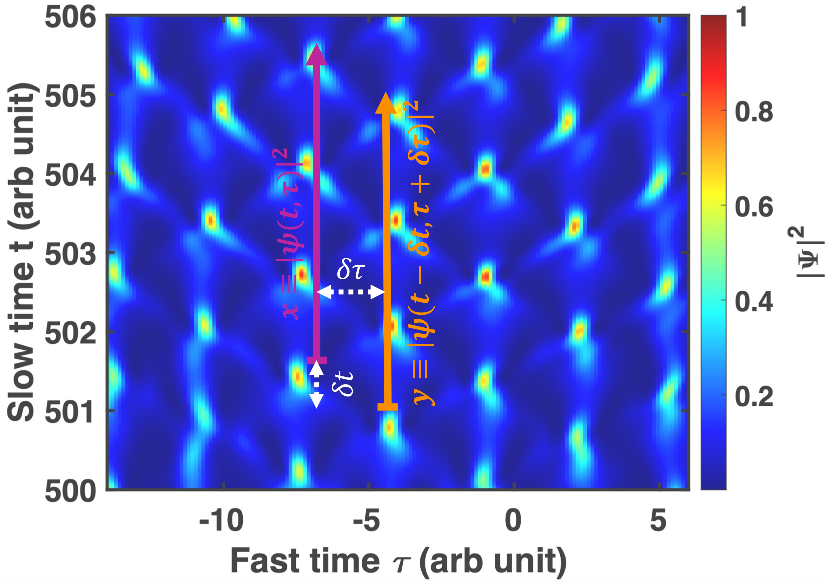


Figure 2.15: Illustration of how the signals are selected to compute the transfer entropy map

The transfer entropy 2D map is presented in Fig. 2.16. At finite roundtrips, it exhibits either a central peak (Pu_i) or double peak (Pr_i) structures. These peaks mean that, on average, any peaks in the evolution carry information from its own past. This information vanishes roundtrip to roundtrip. The dual-peak structure of the Pr_i has the advantage of being easily differentiated compared to the single peak of the Pu_i making precursors the better choice than bursts. Furthermore, the space and round-trip time delays between the peaks of the Pr_i are of the same order as the Fast time periodicity (FTP) and STP, respectively. This approves the choice of the order of magnitude of our subdomains. Each measurement is locally centered at the location of the intensity burst.

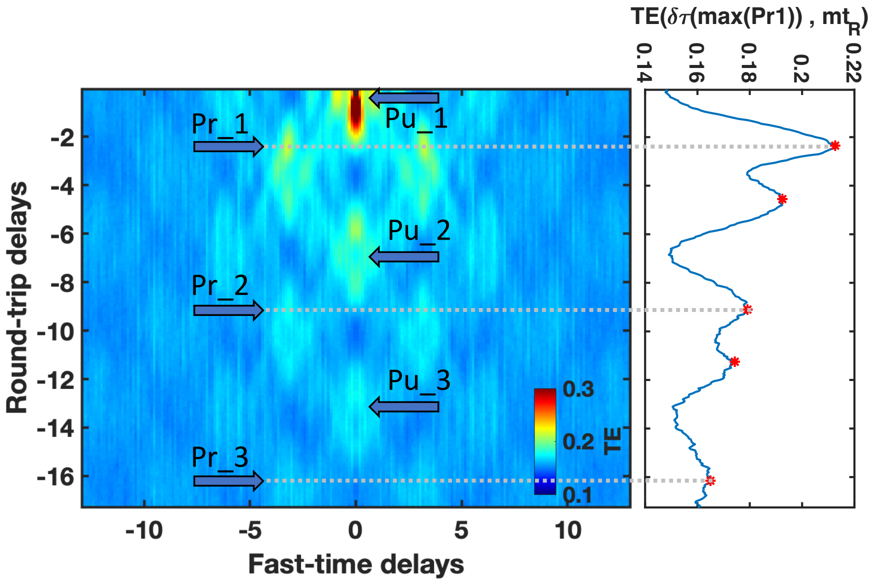


Figure 2.16: The left panel shows the 2D map of the transfer entropy, highlighting the fast-time delays as a function of round-trip delays and highlighting regions with significant information flow, indicated by the Pr_i and Pu_i peaks. The right panel shows the evolution of the transfer entropy at the fast-time delay corresponding to the Pr_1 maximum with respect to the roundtrip delays. The blue solid line represents the variation of the transfer entropy across roundtrips, with symbols (*) marking the peaks at each Pr_i.

2.6.3 Identification of precursor-pulse pairs

These precise measurements from the correlation functions and the TE guide our selection of the subsets in our dataset $|\psi|^2$, where potential precursors are expected to occur. Through a 2D convolution operation of this subset with two Gaussian-like functions, we enhance the visibility of potential precursors by emphasizing local maxima in the convolution output, both spatially and temporally. The convolution operation acts as a filter that emphasizes areas where the characteristics of the precursors are most pronounced, primarily by boosting the signal around local maxima. Local maxima that align with the spatial or temporal boundaries are excluded to ensure the uniqueness and relevance of the detected precursor. From the spatial and temporal coordinates of the identified maxima, we identify a region around these maxima as the precursor by taking a region larger than half the STP in time and also larger than half the FTP in both directions in space. Then precursors are associated with previous pulses occurring at intervals defined by half the STP, effectively pairing each precursor with its corresponding pulse. The subset is then shifted by half the STP, and the procedure is repeated until the end of the dynamic length is reached. In Fig. 2.17, we show an example of this convolution operation done on a subset of $|\psi|^2$.

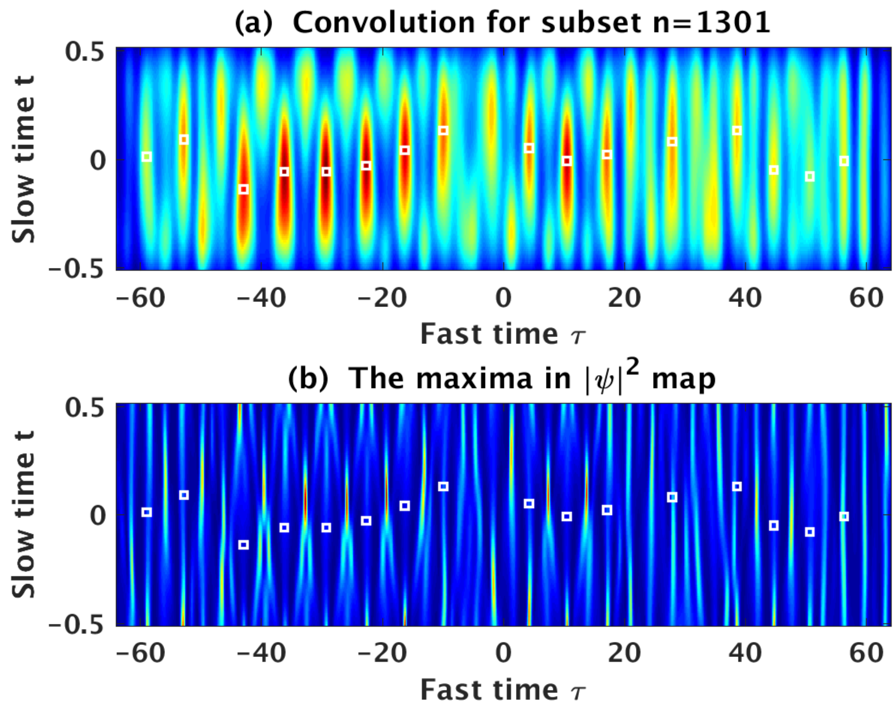


Figure 2.17: Example of the convolution operation and the subsequent identification of maxima in a subset of $|\psi|^2$.

Fig. 2.18 shows a precursor-pulse pair in the spatiotemporal dynamic and the first, second, and third target pulses that we are trying to forecast coming after the precursor by one, three, and five intervals of this measured half-STP corresponding to approximately 3.4, 10, and 17 round-trips, respectively. The FTP and STP are also shown on the map for more clarity. Note that this is a different precursor than the ones shown in Fig. 2.17. In Fig. 2.19, we first show the images corresponding to the seventh maxima shown in Fig. 2.17, we first show the associated pulse of the precursor identified at the location of the maxima, then the precursor itself, then the first, second, and third target pulses. The middle panel represents the zoomed view, and the last panel corresponds to the dataset input and outputs to the RNN.

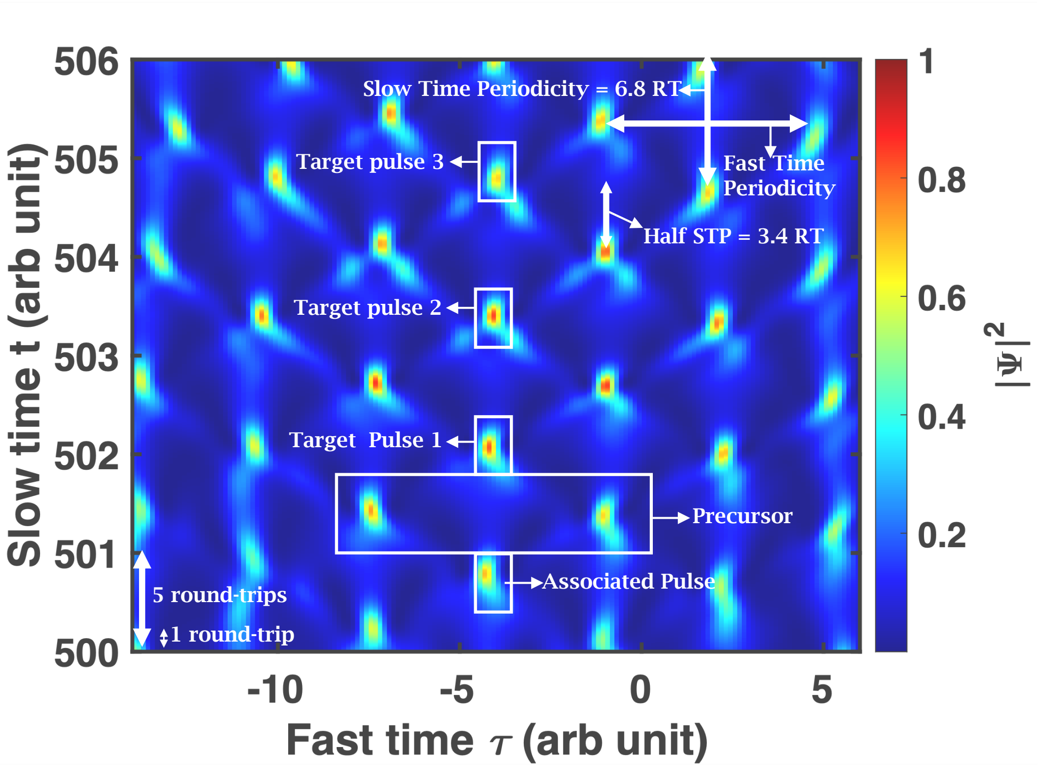


Figure 2.18: Illustration of a precursor-pulse pair in the spatiotemporal dynamic. The first, second, and third target pulses, which are the focus of our prediction, occur after the precursor by one, three, and five intervals of the measured half-STP, corresponding to approximately 3.4, 10, and 17 round-trips, respectively.

2.6.4 Data preparation for Local Forecasting

In total, around 74500 precursors were identified in the spatiotemporal dynamics shown in Fig. 2.2. The dataset comprises images of precursor-pulse

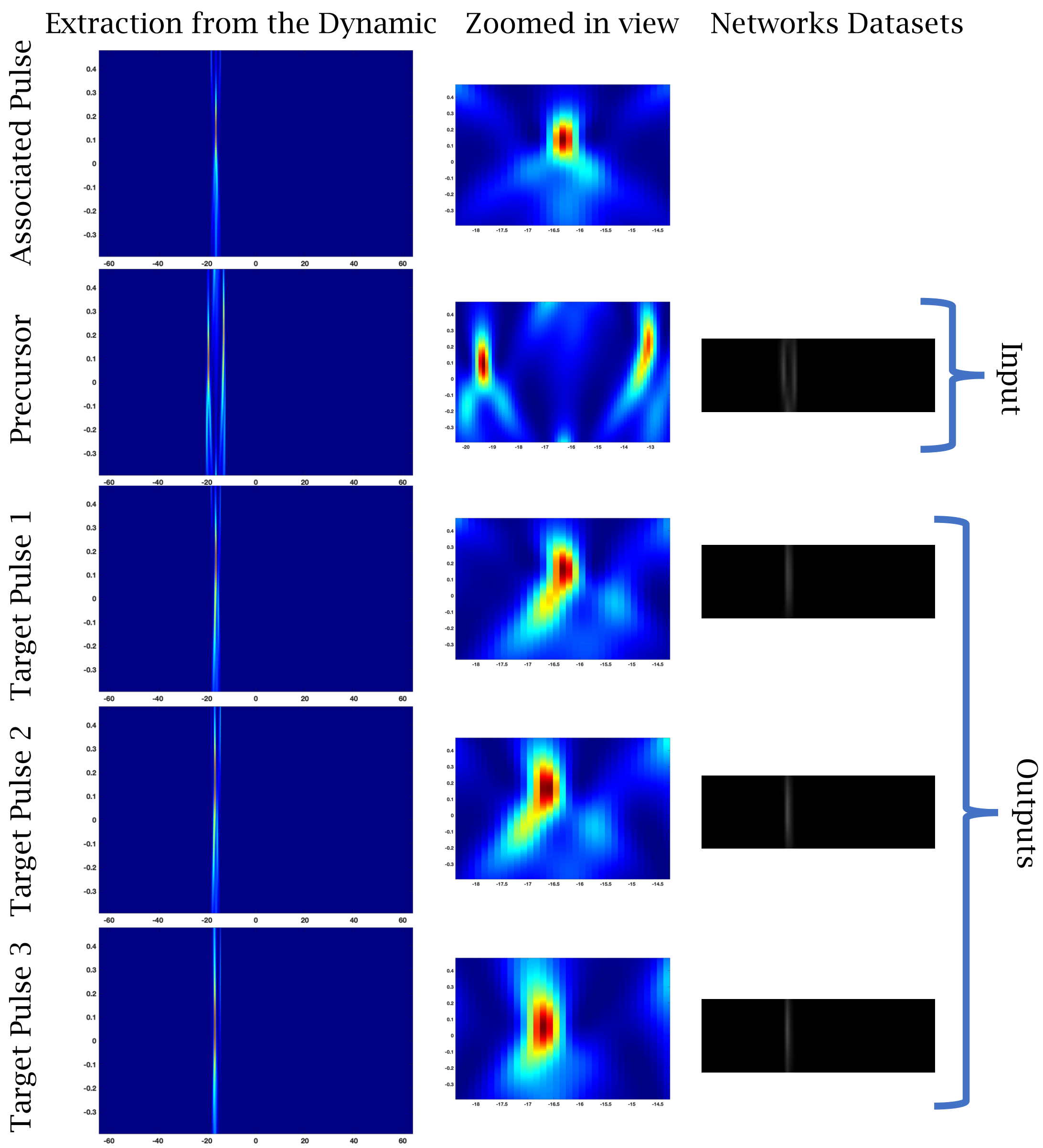


Figure 2.19: Images corresponding to the seventh maxima shown in Fig. 2.17. The associated pulse of the precursor is identified at the location of the maxima, followed by the precursor itself, and the first, second, and third target pulses. The middle panel represents the zoomed view, and the last panel corresponds to the dataset input and outputs to the RNN.

pairs and the corresponding target pulses; an example of the datasets is already shown in the right panel of Fig. 2.19. Precursor images are used as inputs, and target pulse images serve as outputs for training our models. Before splitting the data into training, testing, and validation sets, the dataset is randomly shuffled. This shuffling process is crucial as it ensures that the data is well-mixed, preventing any order-based biases from influencing the training process. Without shuffling, the data might contain unintended patterns or sequences that could lead the model to learn spurious correlations. By ensuring a randomized and well-distributed dataset, we enhance the robustness of our model, leading to more reliable and accurate predictions. The images in the dataset are in grayscale. After shuffling, the data is normalized by scaling the pixel values of the images to a range of $[0, 1]$. This normalization process ensures that all input data falls within the same range, which helps accelerate the training process and improves the model's convergence. Normalizing the data also prevents any feature from dominating due to its scale, ensuring a balanced and fair learning process. By implementing both shuffling and normalization, we ensure that our models are trained on a robust and well-prepared dataset, leading to more reliable and accurate forecasting of extreme events within the chaotic system. The datasets are then divided into 70% for training, 15% for testing, and 15% for validation. The training set is used to train sequence-to-sequence forecasting RNN models, which are designed to utilize the sequential data of identified precursor-pulse pairs to predict future pulses. The validation set is employed to evaluate the model during the training phase, helping to tune the model's hyperparameters and avoid overfitting. The testing set is used for the final assessment of the model's predictive accuracy, providing an unbiased evaluation of its performance on unseen data.

2.6.5 Models Architecture for Local Forecasting

The models used for local forecasting are adapted versions of those employed for full dynamic forecasting discussed in Section 2.5.2. The LSTM model for local forecasting is expanded to include an additional LSTM layer and an extra dropout layer compared to the full dynamic forecasting model. This modification allows it to capture more detailed temporal dependencies within the data and enhance the model's capacity to learn from complex patterns.

The GRU model for local forecasting retains a similar architecture to the one used for full dynamic forecasting, with two GRU layers.

Both models are compiled and fitted using the same methods as those for full dynamic forecasting, including the use of the Huber loss function and the Adam optimizer. We evaluated model performance using the RMSE during training to ensure accurate and reliable predictions. The training process incorporates a learning rate scheduler and early stopping to optimize performance and prevent overfitting.

2.6.6 Hyperparameters tuning

Since there are no predefined hyperparameters for the models we are using, we conducted multiple tests to determine the optimal batch size and number of hidden units. The batch size refers to the number of training examples used in one iteration before updating the model parameters. A smaller batch size allows the model to update more frequently, potentially leading to faster convergence but can also introduce more noise in the updates. Conversely, a larger batch size provides smoother updates, improving the stability of the training process but requires more memory and computational resources.

The number of hidden units, on the other hand, determines the capacity of the model to learn from the data. Hidden units are the neurons in each layer of the model that process the input data and capture patterns within it. A low number of hidden units might lead to underfitting, where the model fails to capture the underlying patterns of the data, resulting in poor performance. Conversely, a high number of hidden units can lead to overfitting, where the model learns the noise in the training data, leading to poor generalization to new, unseen data. Therefore, finding a balance between too few and too many hidden units is crucial for achieving optimal model performance. The best combination of batch size and hidden units was identified, ensuring that the model can effectively learn from the data while maintaining good generalization capabilities.

Through extensive experimentation, we found that a configuration of 256 hidden units and a batch size of 128 consistently provided the best performance across different system complexities and prediction horizons. This configuration balanced model complexity and computational efficiency, yielding reliable and accurate predictions. Some configurations with higher hidden units and batch sizes showed marginally better performance, but the increased computational cost made them less practical.

2.6.7 Evaluating Model Accuracy for Local Forecasting

The results of our experiments, which involved varying system complexities, round-trip delays, hidden units, and batch sizes, provide significant insights into the optimal configurations of the LSTM and GRU networks.

Pearson correlation

For the system with $I_s = 2.2$:

- **Target Pulse 1:** The PC for LSTM was 87% for maximum and 86% for standard deviation, while for GRU it was 80% for maximum and 78% for standard deviation (Fig. 2.20a).
- **Target Pulse 2:** The PC for LSTM was 81% for maximum and 78% for standard deviation, while for GRU it was 74% for maximum and 68% for standard deviation (Fig. 2.20a).
- **Target Pulse 3:** The PC for LSTM was 71% for maximum and 70% for standard deviation, while for GRU it was 59% for maximum and 58% for standard deviation (Fig. 2.20a).

For the system with $I_s = 2.8$:

- **Target Pulse 1:** Despite the increased complexity of the system compared to $I_s = 2.2$, the setup with 256 hidden units and a batch size of 128 provided the best predictive performance for both LSTM and GRU models. However, it was observed that the accuracy decreased when working with the system of $I_s = 2.8$, suggesting that the increased complexity of the system negatively affects the performance of the networks. The PC for LSTM was 67% for maximum and

71% for standard deviation, while for GRU it was 65% for maximum and 69% for standard deviation (Fig. 2.20b).

- **Target Pulse 2:** The PC for LSTM was 49% for maximum and 53% for standard deviation, while for GRU it was 49% for maximum and 51% for standard deviation. While configurations with 512 hidden units and a batch size of 256, as well as 1024 hidden units and a batch size of 128, showed slightly better performance, the improvement was minimal compared to the significant increase in computational cost (Fig. 2.20b).

- **Target Pulse 3:** The PC for LSTM was 40% for maximum and 43% for standard deviation, while for GRU it was 41% for maximum and 42% for standard deviation (Fig. 2.20b).

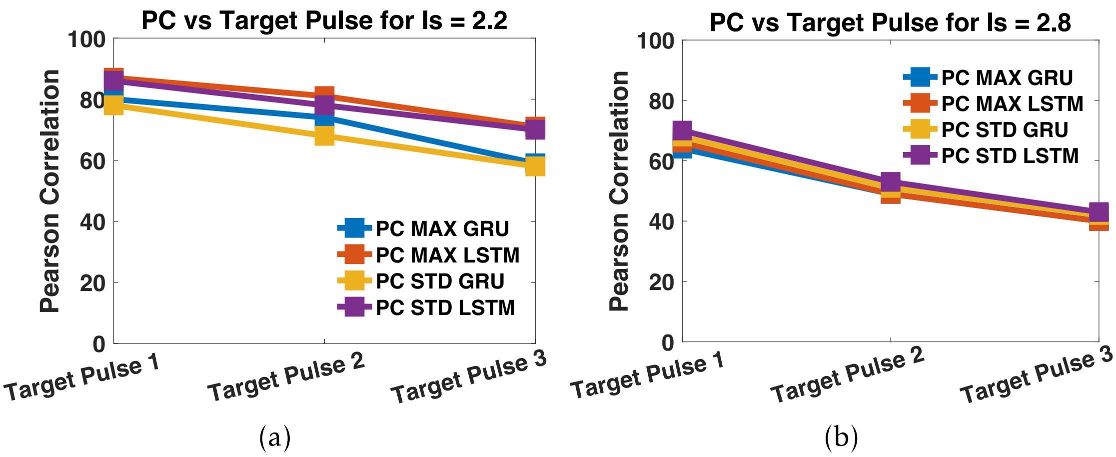


Figure 2.20: Comparison of PC values across different target pulses for LSTM and GRU models with optimal configurations for systems with (a) $I_s = 2.2$ and (b) $I_s = 2.8$.

The findings from our experiments highlight several key insights into the performance and configuration of the LSTM and GRU networks across different system complexities and round-trip delays. For simpler systems with $I_s = 2.2$, the configuration of 256 hidden units and a batch size of 128 generally provided the best performance across all target pulses. For more complex systems with $I_s = 2.8$, the same configuration of 256 hidden units and a batch size of 128 remained optimal, although the performance decreased compared to the simpler system. As the round-trip delays increased, the performance of the models generally decreased, reflecting the increased difficulty in predicting further into the future. Despite this, the chosen configurations managed to maintain relatively high performance, especially for the simpler system. In comparison, the LSTM models consistently performed well with the optimal configuration across different system complexities and round-trip delays. The GRU models, while showing some improvement with higher configurations, did not justify the additional computational cost. These findings underscore the importance of carefully tuning hyperparameters to balance model complexity and computational efficiency, particularly as system complexity and prediction horizon increase. The identified optimal configurations provide a robust foundation for effectively predicting extreme events in spatiotemporal chaos.

Other than the PC, we used the Symmetric Mean Absolute Percentage Error (sMAPE) and the Median Symmetric Accuracy (ζ) metrics to evaluate the model performance.

Symmetric Mean Absolute Percentage Error and Median Symmetric Accuracy Histograms

The Symmetric Mean Absolute Percentage Error (sMAPE) and Median Symmetric Accuracy (ζ) are robust metrics for evaluating the performance of our LSTM and GRU models. sMAPE measures the accuracy based on percentage errors, making it scale-independent and useful for comparing forecast performances across datasets [184, 185]. sMAPE is given by the following formula:

$$\text{sMAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|F_i - A_i|}{(|A_i| + |F_i|)/2} \tag{2.4}$$

where F_i and A_i are the forecasted and actual values, respectively.
ζ measures the median relative accuracy and is particularly resilient to skewed distributions [185, 186], calculated as:

$$\zeta = 100 \times \left(\exp\left(\text{median}\left(\left| \ln\left(\frac{F_i}{A_i} \right) \right| \right) \right) - 1 \right) \tag{2.5}$$

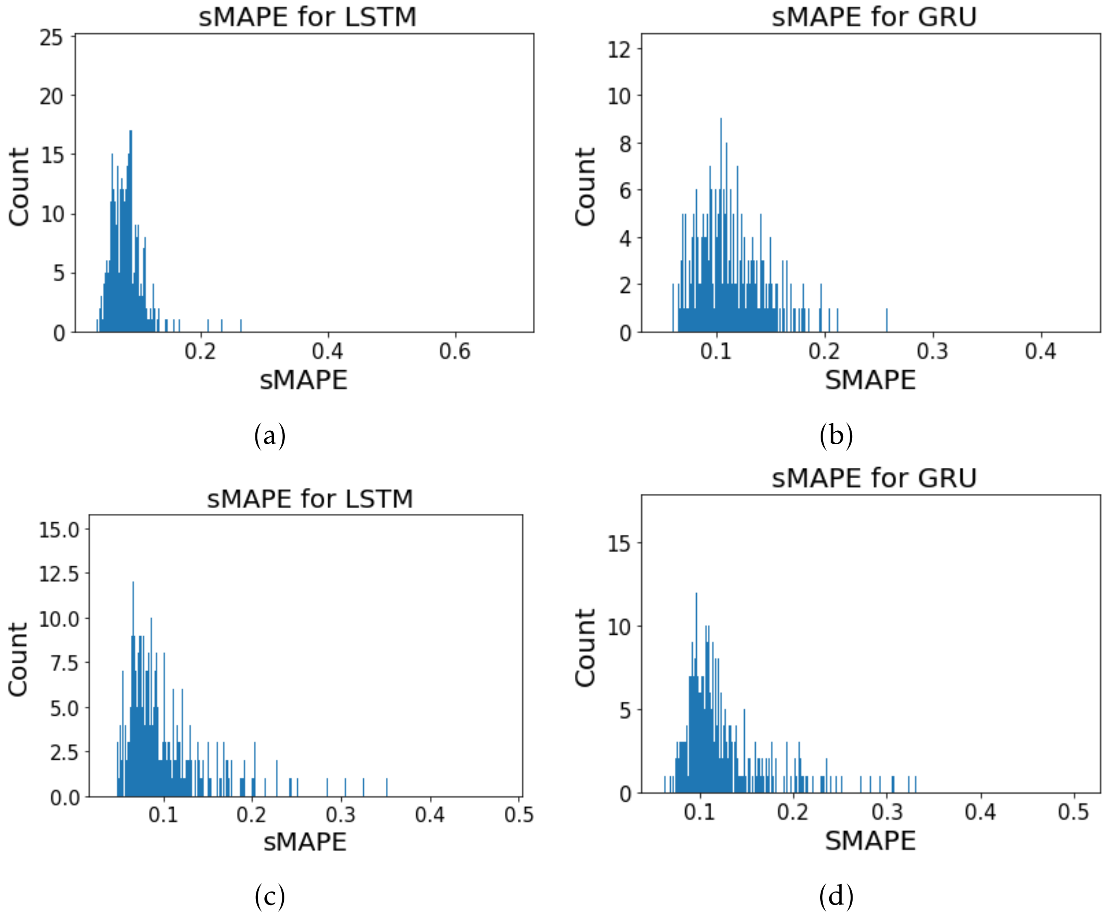


Figure 2.21: sMAPE histograms for the prediction of the first target pulse using (a-b) LSTM and (c-d) GRU models for systems with $I_s = 2.2$ and $I_s = 2.8$, respectively.

The histograms of sMAPE and ζ for the prediction of the first target pulse using the LSTM and GRU models for systems with $I_s = 2.2$ and $I_s = 2.8$ are shown in Figs. 2.21 and 2.22. The histograms demonstrate clear distinctions in model performance. The LSTM model consistently shows lower sMAPE values compared to the GRU model, indicating higher accuracy in predicting the target pulses. Similarly, the ζ histograms reveal that the LSTM model has a tighter distribution of errors, suggesting more reliable predictions. This observation is consistent with the PC values discussed earlier, where the LSTM model exhibited superior performance.

The comparison between LSTM and GRU models across different system

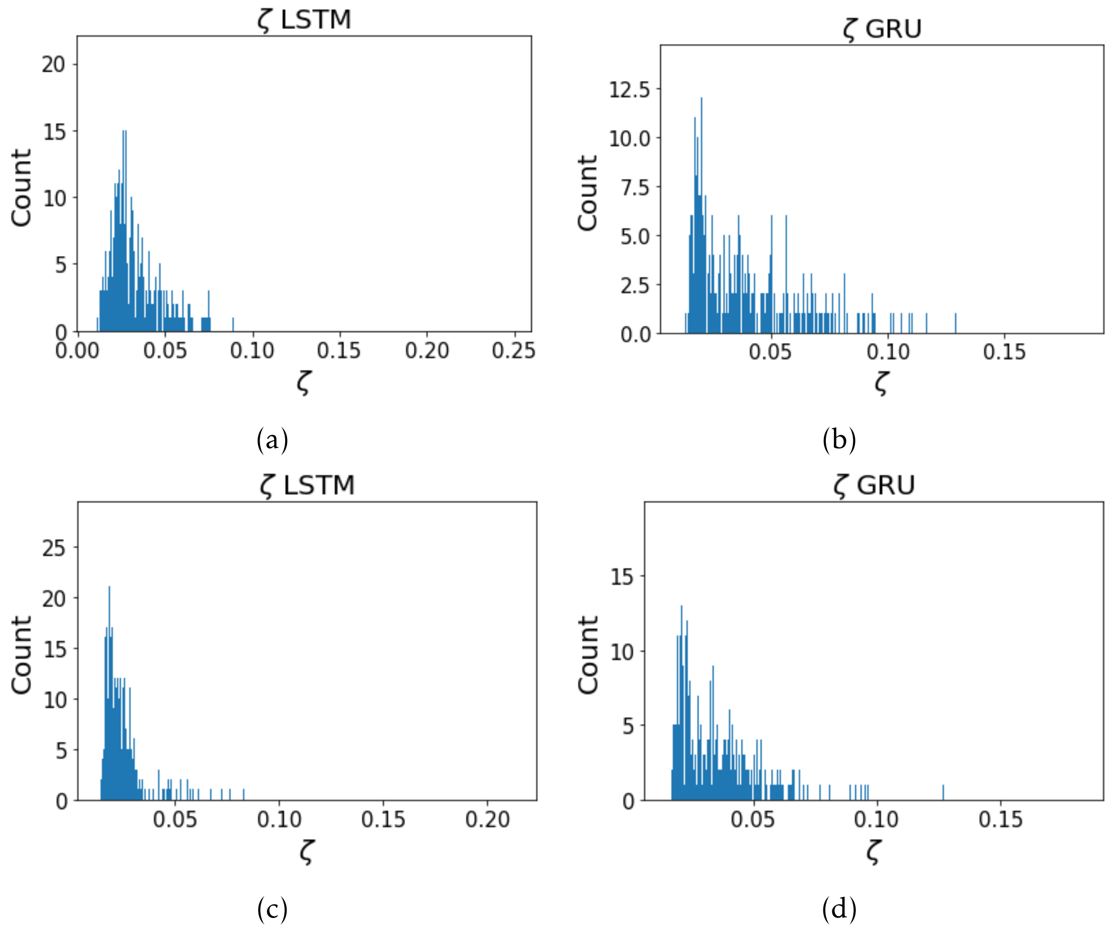


Figure 2.22: ζ histograms for the prediction of the first target pulse using (a-b) LSTM and (c-d) GRU models for systems with $I_s = 2.2$ and $I_s = 2.8$, respectively.

complexities ($I_s = 2.2$ and $I_s = 2.8$) reveals that the LSTM model generally outperforms the GRU model. This performance gap is more pronounced in more complex systems ($I_s = 2.8$), where the LSTM model maintains higher accuracy and stability.

In conclusion, the LSTM model is superior in both accuracy and reliability for local forecasting in systems with varying complexities. The histograms of sMAPE and ζ metrics corroborate these findings, further establishing the LSTM model as the preferred choice for predicting extreme events in spatiotemporal chaos.

2.6.8 Predicted Pulses

A visual comparison of the predicted pulses using the LSTM and GRU models for systems with $I_s = 2.2$ and $I_s = 2.8$ is shown in Fig. 2.23. From this figure, it is evident that the pulses predicted by the LSTM model appear more accurate and well-defined compared to those predicted by the GRU model. This observation is consistent with the PC performance of both models, further highlighting the superior performance of the LSTM model in predicting the pulse shape.

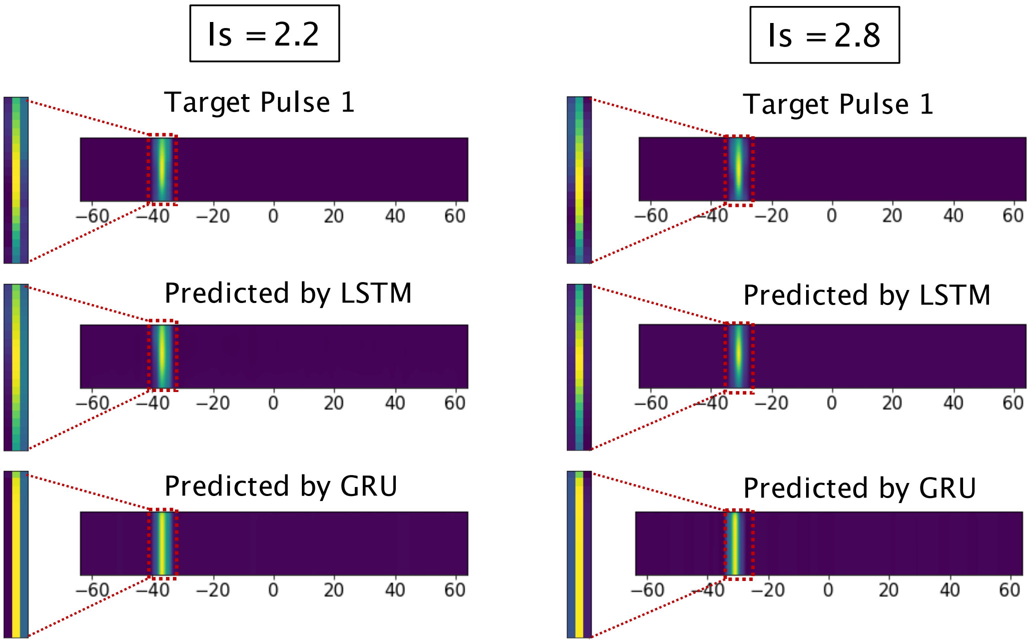


Figure 2.23: Comparison of the first target pulse and the predicted pulses using the LSTM and GRU models for systems with $I_s = 2.2$ and $I_s = 2.8$.

To conclude this section, our local forecasting method, which uses precursor identification through statistical analysis and information theory, has proven effective in predicting the specific future events within spatiotemporal chaos. This method successfully identifies critical precursors that precede significant events, allowing for accurate forecasting of upcoming pulses. Specifically, we have demonstrated that local forecasting can predict these pulses up to 17 cavity round trips. This approach not only enhances the reliability of predictions in chaotic systems but also contributes to a deeper understanding of the underlying dynamics of spatiotemporal chaos.

2.7 Conclusion

In this chapter, we examined the predictability of spatiotemporal chaos dynamics within optical fiber ring resonators through the application of advanced machine learning techniques. Our primary focus was on the implementation and evaluation of two forecasting methodologies: full dynamic forecasting and local forecasting based on precursor identification.

Initially, we employed Recurrent Neural Networks (RNN), specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, to forecast the complete spatiotemporal dynamics of the system. This approach yielded several significant insights. The models demonstrated robust predictive capabilities for less complex systems characterized by smaller sizes and lower pumping values. However, as the system complexity increased, the prediction accuracy notably declined, particularly for long-term forecasts. The Pearson correlation (PC) metric dropped below 50% beyond a critical threshold, indicating a substantial limitation in the models' ability to reliably predict future states in high-dimensional chaotic systems.

Recognizing the limitations of the full dynamic forecasting approach, we introduced an alternative method termed local forecasting. This method focused on predicting the occurrence of extreme events based on identified precursors, providing a more targeted and precise forecasting approach. Through statistical analysis and information theory, we successfully identified precursors that enabled the prediction of future extreme events. The local forecasting method proved effective in predicting upcoming pulses up to 17 cavity round trips in advance, highlighting its potential in managing and mitigating the impacts of extreme events in chaotic systems. Throughout our study, we also conducted a comparative analysis of the LSTM and GRU models. Both models exhibited similar trends in predictive accuracy, with LSTM marginally outperforming GRU in capturing long-term dependencies due to its more complex memory management mechanisms. However, the differences between the two models were minimal, and both faced challenges in maintaining accuracy for highly complex systems over extended periods.

In summary, our work underscores the potential and limitations of using machine learning techniques for predicting spatiotemporal chaos dynamics. While full dynamic forecasting provides a comprehensive overview, local forecasting offers a more precise and actionable approach, particularly in predicting extreme events. Future research should aim to enhance model robustness and explore hybrid approaches to improve long-term forecasting accuracy in high-dimensional chaotic systems.

Chapter 3

Characterization of Spatiotemporal Chaos and Turbulence induced by Solitons in Optical Systems

Outline of the current chapter

This chapter marks a transition from the monostable regime within an optical fiber ring resonator explored in Chapter 2, to an investigation of the dynamics within the bistable regime, where we aim to delve deeper into the complex dynamics that emerge under these conditions. Our primary objective in this chapter is to comprehensively understand these dynamics and to demonstrate the turbulent-like behavior that arises in this system.

3.1 Introduction

Starting with a soliton as the initial condition at the onset of the bistable regime and progressively increasing the pump power, we can trace the evolution of the system's dynamics. This process, well-documented in the literature, shows the soliton undergoing a Hopf bifurcation, followed by period doubling, ultimately giving rise to spatiotemporal chaotic dynamic, eventually leading to fully developed turbulence. Turbulence, a fundamental concept in fluid dynamics, is known for its complex spatiotemporal behavior, and its application has extended to other systems out of equilibrium, including nonlinear optics, chemical reactions, active matter, and even economics.

Our main goal in this chapter is to characterize this chaotic dynamic. We begin by identifying key characteristic lengths of the system, including two-point correlation lengths in time and space, the Lyapunov dimension length [102], and the inter-pulse distance. We will examine the power spectra of the phase and amplitude in both space and time. This analysis will provide a statistical foundation for identifying and characterizing the turbulent-like behavior. Specifically, we aim to show that phase turbulence within the bistable regime is distinguished by power spectra with an exponent of -2, indicating a self-similar structure across different scales. However, the power-law spectra and these characteristic lengths do not fully capture the system's intricate behavior [98].

To better understand the fluctuations and determine whether they exhibit characteristic scales, we then analyze higher-order correlations, also known as structure functions [98]. Through this approach, we reveal that the system exhibits intermittent behavior. Similar behavior has been observed in other systems, such as the work on bi-turbulence in fingerprint pattern formations, where the transition from self-similarity to intermittency was similarly noted [187].

Unlike in systems governed by modulation instability, where extreme events are strongly nonlinear and dominated by turbulence, the extreme events observed in this system are governed by phase dynamics. These dynamics follow linear wave behavior and are further influenced by the appearance of defects. This distinction emphasizes that while turbulence typically drives nonlinear extreme events, here, the extreme events emerge from a combination of linear wave propagation and the formation of phase defects, marking a different mechanism of instability and event formation.

Following this, we examine the dispersion relation to investigate how linear waves influence the system's dynamics, even within this predominantly turbulent regime. The dispersion relation provides further insight into the interaction between linear wave instabilities and the nonlinear chaotic behavior observed in the system.

3.2 Numerical Simulation Setup

Unlike the previous chapter, where the Lugiato-Lefever Equation (LLE) was normalized by the cavity loss parameter α, in this chapter, we retain α in its original form. The LLE in this form is given by:

$$\frac{\partial\psi(t,\tau)}{\partial t} = S - (\alpha + i\delta)\psi + i\frac{\partial^2\psi}{\partial\tau^2} + i\mid\psi\mid^2\psi \tag{3.1}$$

By setting the parameters as $\delta = 1$, $\alpha = 0.16$, we find that the normalized detuning parameter $\Delta = \frac{\delta}{\alpha} = 6.25$, placing the system in the bistable regime, as

$\Delta > \sqrt{3}$. The steady-state phase ϕ_s is given by:

$$\phi_s = \tan^{-1}\left(\frac{I_s - \Delta}{\alpha}\right) \tag{3.2}$$

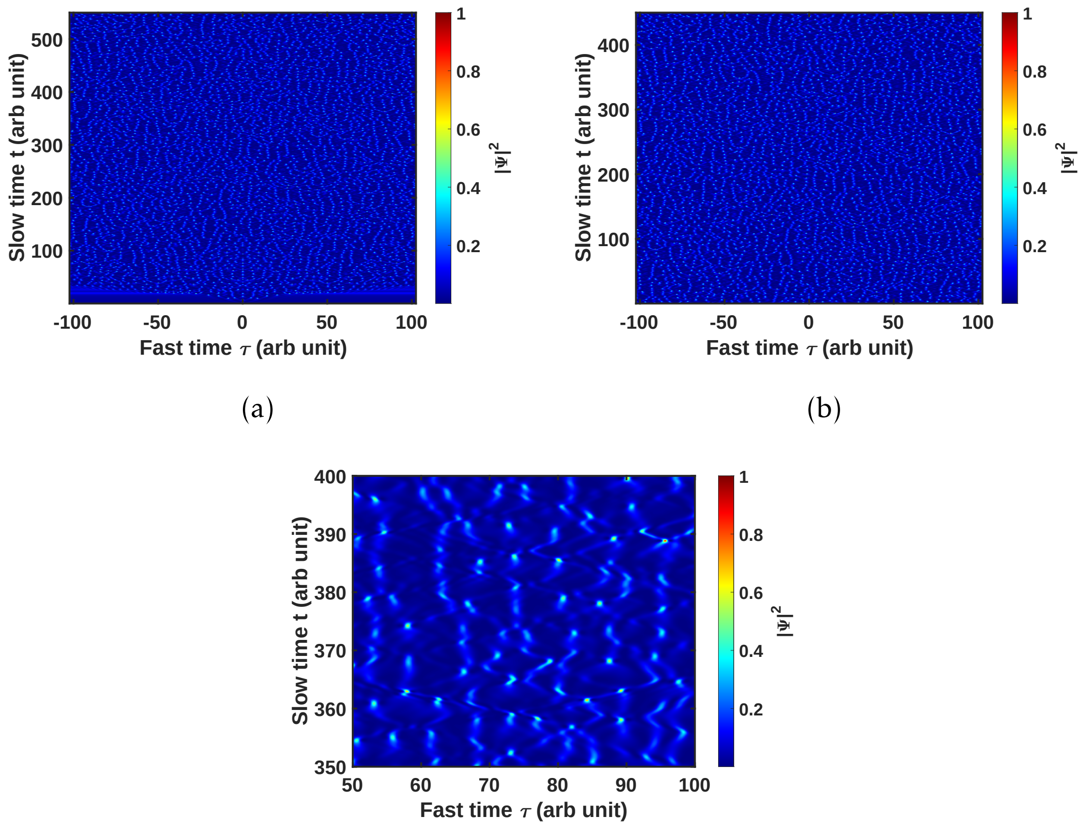
The homogeneous and steady-state solutions of Eq. 3.1 follow the cubic equation:

$$|S|^2 = I_s[\alpha^2 + (I_s - \Delta)^2] \tag{3.3}$$

These formulations establish the groundwork for our detailed investigation into the bistable regime, enabling us to understand and characterize the complex behaviors and dynamics that emerge under these conditions. We numerically solved the LLE for the parameters mentioned above, with a pump strength $S = 0.4$, using a spatial grid with $N_p = 4096$. The simulation spans a total duration from $t = 0$ to $t = 550$, corresponding to 3437 cavity round trips. Initially, the system starts with a coherent soliton as the initial condition. This soliton undergoes dynamic evolution, transitioning from a stable state to a self-pulsating state due to the interplay between nonlinearity and cavity parameters. As the system evolves further, these pulsations become increasingly irregular and complex, ultimately leading to a state of spatiotemporal chaos characterized by unpredictable and irregular oscillations in both space and time. The space-time map (Fig 3.1a) illustrates this transition from a stable soliton to spatiotemporal chaos. To focus solely on the chaotic dynamics, the transient phase was removed, as shown in Fig. 3.1b, with a zoomed-in view provided in Fig. 3.1c.

3.2.1 Phase Dynamics: Wrapping and Unwrapping

Previously, we focused on analyzing the intensity dynamics within the bistable regime. Building upon this foundation, we investigated the phase behavior of the system. However, a common challenge in phase analysis is the phenomenon of phase wrapping. Phase wrapping occurs when phase measurements are constrained to a specific range, typically between $-\pi$ and π (or 0 to 2π). In many measurement systems, the phase is measured modulo 2π, meaning that once the phase exceeds these bounds, it wraps back around. This is similar



(a) (b) (c)

Figure 3.1: Space-time map starting from a soliton into spatiotemporal chaos (a) with, (b) without the transitory phase, and (c) a zoomed map for clarity.

to how a clock wraps back around every 12 hours. As a result, the true phase might be indistinguishable from its wrapped version, leading to ambiguity. To overcome this challenge, we employ phase unwrapping, a process that resolves this 2π ambiguity and reconstructs the actual phase values from the wrapped phase. The goal of phase unwrapping is to determine the number of 2π jumps needed to restore the actual, continuous phase values. For this purpose, we used a method based on the reference [188], which performs deterministic phase unwrapping. Deterministic methods for phase unwrapping typically follow the phase gradients and make decisions on adding or subtracting 2π based on predefined rules, aiming to maintain phase continuity. These methods can vary in complexity and robustness, especially in the presence of noise, which can complicate phase unwrapping.

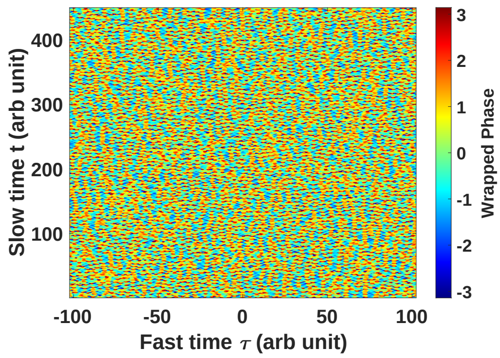
In Fig. 3.2, we present the space-time maps of the wrapped and unwrapped phase of the field in the cavity for $S = 0.4$, which reveal different patterns of phase variability over time and space. Figs. 3.2a and 3.2b show the wrapped phase and a zoomed-in view, where the phase values are constrained within a specific range, typically between $-\pi$ and π. This results in a periodic-like pattern where the phase abruptly jumps from one end of the range to the other, giving a striped appearance. In contrast, Figs. 3.2c and 3.2d represent the unwrapped phase and a zoomed-in view, where the phase values are allowed to exceed the typical $-\pi$ to π range and can increase or decrease without bound. This results in a smoother transition between phase values without the abrupt jumps, giving a more gradient-like pattern.

The key difference between the two representations is that the wrapped phase presents discontinuities due to the wrapping effect, while the unwrapped phase shows the continuous evolution of the phase over time and space. This unwrapping is useful as it removes the artificial periodicity imposed by wrapping.

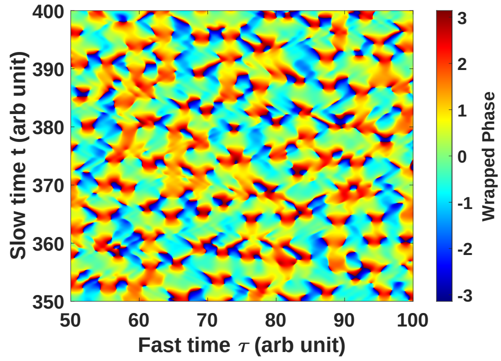
3.2.2 Fourier Analysis of Spatial and Temporal Spectra

To gain deeper insights into the dynamics of our system in the bistable regime, it is essential to analyze not only the intensity and phase behavior but also the spatial and temporal spectra of the phase and amplitude. By performing Fourier analysis on both the phase (wrapped and unwrapped) and amplitude, we can better understand how the spectra is distributed across different frequencies, which is crucial for identifying scaling laws and the turbulent-like behavior in the system.

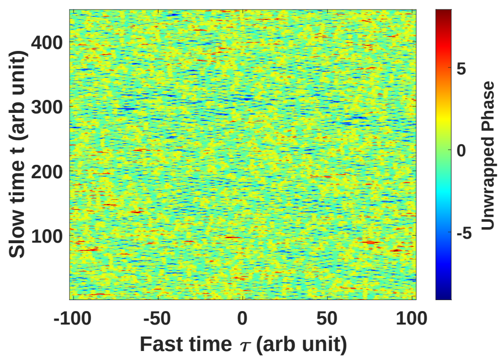
Fig. 3.3a (Left panel) shows the temporal spectra for both the wrapped and unwrapped phases. In both cases, the spectra exhibit a power-law decay with an exponent close to -2, indicating scale-invariance over a broad range of temporal scales. This behavior suggests self-similarity in the phase dynamics and supports the hypothesis that the phase turbulence in the system follows a classical turbulent cascade [37].



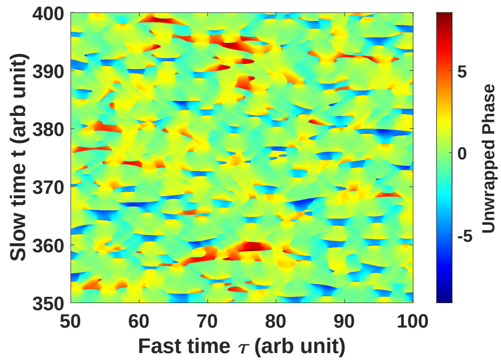
(a)



(b)



(c)



(d)

Figure 3.2: Space-time maps of (a) the wrapped phase and (b) a zoomed-in view, and (c) the unwrapped phase and (d) a zoomed-in view.

On the Right panel of Fig. 3.3a, the temporal spectrum of the amplitude is presented. Unlike the phase, the amplitude does not follow a simple power-law decay. Instead, the spectrum decays more steeply, with an exponent of approximately -5.8, indicating a much more rapid loss of power at higher frequencies. This deviation from the power-law behavior seen in the phase spectra suggests that the amplitude is governed by different dynamics, likely influenced by nonlinear interactions and complex structures such as solitons within the turbulent regime.

The spatial spectra are analyzed in Fig. 3.3b. The Left panel displays the spatial spectra for the wrapped and unwrapped phases, both of which exhibit a power-law decay with an exponent close to -2. This again suggests scale invariance in the phase dynamics across spatial scales, reflecting the presence of spatial turbulence. The self-similarity observed here is consistent with the findings in the temporal spectra.

The Right panel of Fig. 3.3b presents the spatial spectrum of the amplitude, which does not follow a clear power-law. Instead, it shows a steep and complex decay. This steep decay may indicate the influence of solitonic structures that are present but highly disturbed by the turbulent dynamics.

Fig. 3.4 provides the semi-logarithmic representation of the spatial spectrum of the amplitude. The quasi-triangular shape observed in this plot is a signature of soliton-like structures within the system. This behavior is expected in the regime of frequency combs, where discrete spectral lines, or "teeth," are normally visible. However, due to the high level of turbulence in the system, these comb lines have merged into a continuous spectrum. This indicates that while the system retains some solitonic features, the chaotic nature of the dynamics obscures the regular structure of the frequency comb.

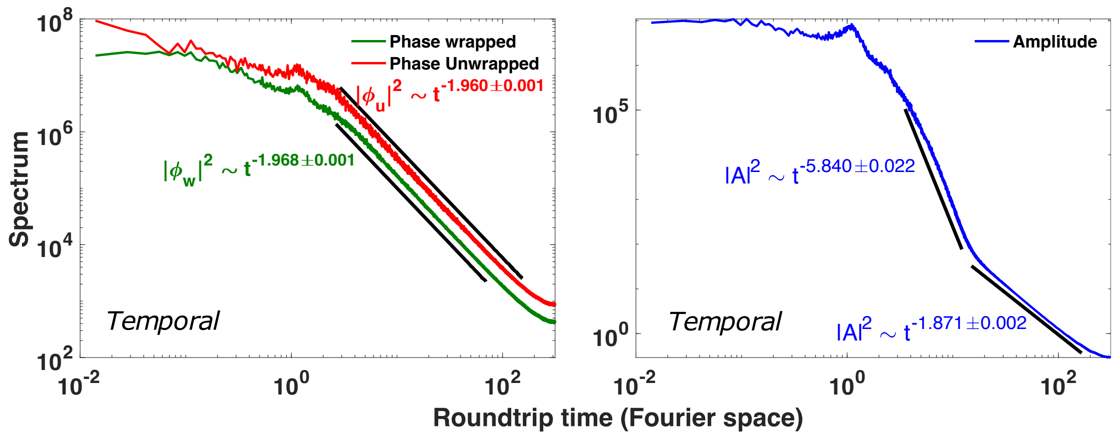
In the upcoming sections, we will further investigate the dynamics of the system using multiple analytical and numerical methods. This will allow us to better understand the complex behaviors observed in the phase and amplitude spectra.

3.3 Two-Points Correlation Functions for Spatial and Temporal Dynamics

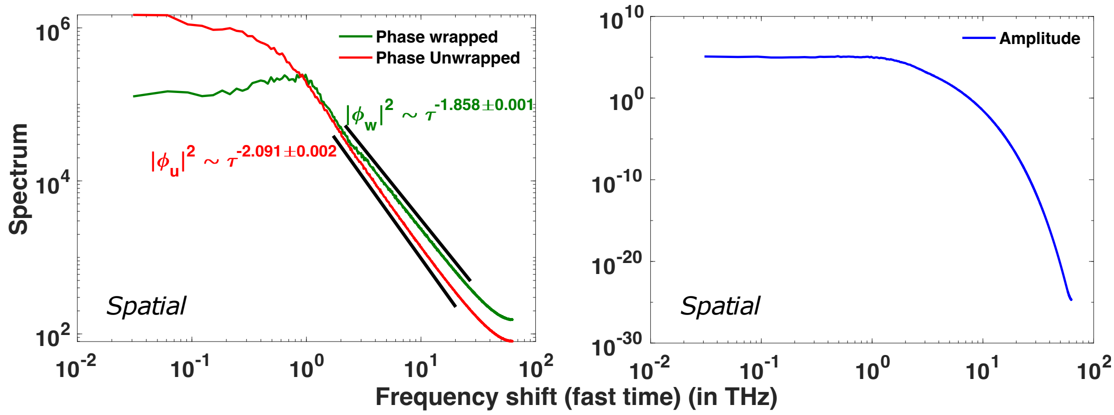
In our investigation of the system's complex evolution, we examine the probability that two locations separated by $\delta\tau$ (spatial separation) and δt (temporal separation) respectively behave coherently. This coherence is quantified using the two-points correlation functions $C(\delta\tau)$ (Eq. 3.4) and $C(\delta t)$ (Eq. 3.5):

$$C(\delta\tau) = \langle(\psi(\delta\tau + \tau', t) - \langle\psi\rangle)(\psi(\tau', t) - \langle\psi\rangle)\rangle. \tag{3.4}$$

$$C(\delta t) = \langle(\psi(\tau, \delta t + t') - \langle\psi\rangle)(\psi(\tau, t') - \langle\psi\rangle)\rangle. \tag{3.5}$$



(a)



(b)

Figure 3.3: Temporal spectra for (a) (Left panel) the wrapped and unwrapped phase and (Right panel) the amplitude, Spatial spectra for (b) (Left panel) the wrapped and unwrapped phase and (Right panel) the amplitude, all are in a log scale.

Here, the brackets $\langle\cdot\rangle$ stand for the average process. $C(\delta\tau)$ and $C(\delta t)$ are the equal-time two-points correlation functions spatially and temporally, respectively.

Directly determining $C(\delta\tau)$ and $C(\Delta t)$ can be computationally intensive. However, by employing the Wiener-Khinchin theorem [182, 183], we can compute these functions more efficiently. This involves time-averaging the Fourier spectra and then taking the inverse Fourier transform of their squared magnitudes.

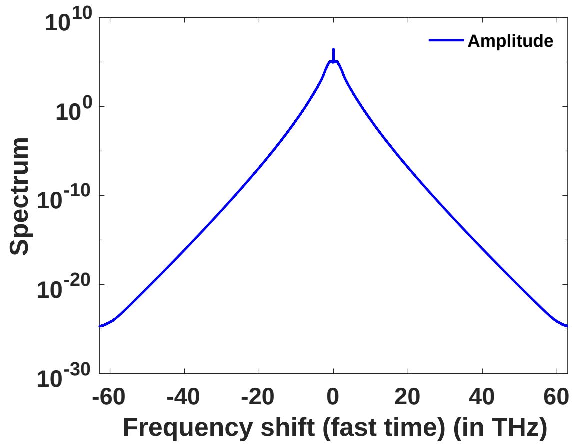


Figure 3.4: Spatial spectra for the amplitude in the semi-log scale

In the context of spatiotemporal chaos, another important quantity is the Kaplan-Yorke dimension D_{KY}, which estimates the fractal dimension of the attractor in phase space. It is given by:

$$D_{KY} = p + \frac{\sum_{i=1}^{p} \lambda_i}{\left|\lambda_{p+1}\right|}, \tag{3.6}$$

where p is the largest integer that satisfies $\sum_{i=1}^{p} \lambda_i > 0$. Here λ_i are the Lyapunov exponents, which measure the rates of separation of infinitesimally close trajectories in phase space. D_{KY} may change linearly with the volume of the system. That is, for a 1D system, $D_{KY} = \xi_\delta^{-1} \Delta T$ where ΔT is the temporal extension of the system and ξ_δ represents the Lyapunov dimension density of the system for a fixed value of the control parameter. This quantity gives an estimation of the extension of the dynamically independent subsystems. Fig. 3.5 taken from article [102] shows that $\xi_\delta = 1/1.73 = 0.57$, note that all quantities in this formulation are expressed in normalized units.

The plot of the spatial two-points correlation function in (Fig. 3.6(a)) shows a very rapid decrease, with the correlation plot having the same shape as a pulse profile. This indicates that outside the pulse, the probability of coherent evolution is nearly zero. This is expected in a turbulent system where coherence

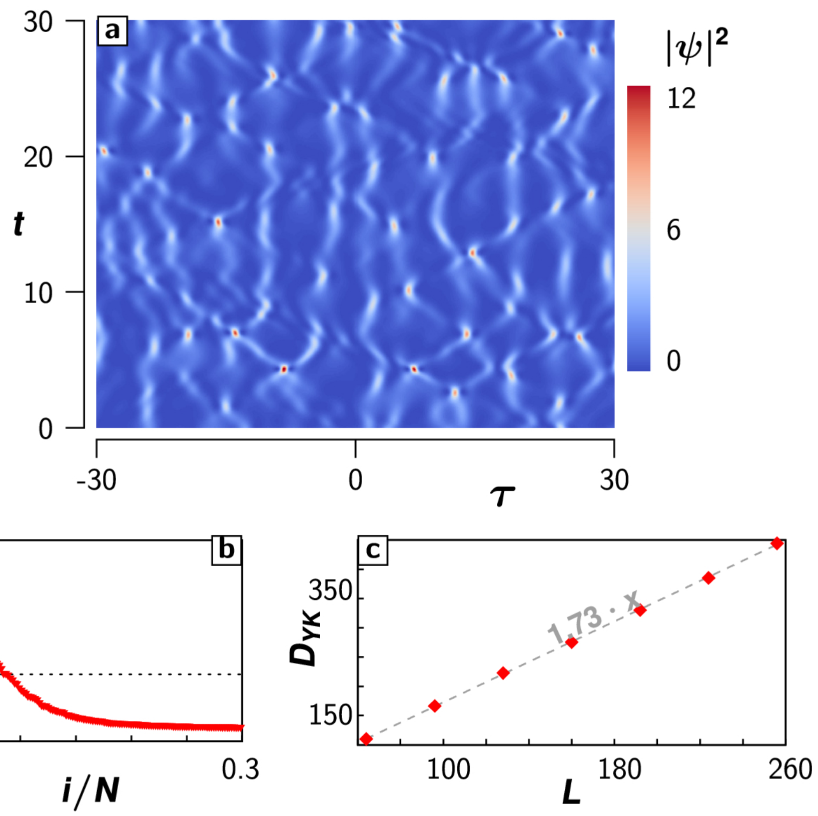


Figure 3.5: Spatiotemporal chaos. (a) $\tau - t$ map shows a complex spatiotemporal behavior obtained by numerical simulation of Eq. (1) with $\alpha = 0.16$, $\delta = 1$, and $S^2 = 0.16$ with 512 grid points. (b) Corresponding Lyapunov spectrum, and (c) York-Kaplan dimension as a function of the system size L is indicated by the diamond red points. $L = 512\Delta\tau$ with $\Delta\tau$ as the step-size integration. The linear growth of D_{KY} dimension is fitted by a slope of 1.73, as shown by the gray dashed line. Figure taken from [102].

is typically confined to small regions. Fig. 3.6(b) provides a zoomed-in view of the central part of Fig. 3.6(a), emphasizing the rapid decay to zero at shorter fast time delays. The best way to define the correlation length spatially is to consider the first zero crossing, which in this case $\xi_2 = 3.1$. Additionally, Fig. 3.6(c) presents the two-points correlation function with the fast time delay divided by the Lyapunov dimension density, highlighting the correlation behavior in normalized units. The correlation length in this normalized form is $\xi_2 = 5.28\xi_\delta$.

To better understand the behavior of the correlation function beyond the

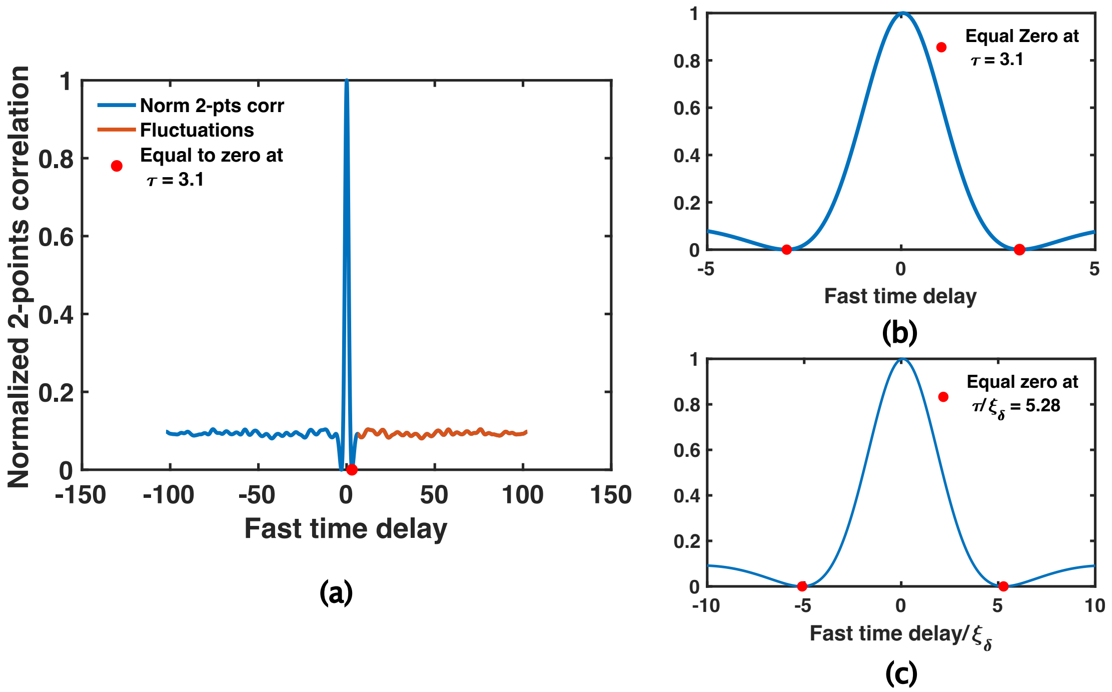


Figure 3.6: Two-points correlation function in space: (a) Overall view showing the rapid decrease. (b) Zoomed-in view of the central part of (a), highlighting the rapid decay to zero. (c) Two-points correlation function with the fast time delay divided by the Lyapunov dimension density.

central peak (highlighted in red in Fig. 3.6(a)), we analyze the fluctuations in this region. These fluctuations provide valuable insights into the coherence properties and the turbulent dynamics of the system. We isolate the right side of the central peak of the correlation function and apply an absolute squared Fourier transform. This process involves calculating the Fourier transform of the selected segment and then shifting the spectrum to center the zero frequency. This approach helps us focus on the specific frequency components that characterize the fluctuations. The resulting frequency-domain representation, shown in Fig. 3.7, provides the power spectrum of the selected segment. The log scale plot of this frequency spectrum reveals a power-law decay with an exponent of -1.82 which is close to the universal exponent -2, a hallmark of phase turbulence. This analysis shows that the phase dynamic is different than the intensity dynamic and that the energy cascade of the intensity comes from the fluctuations around the soliton.

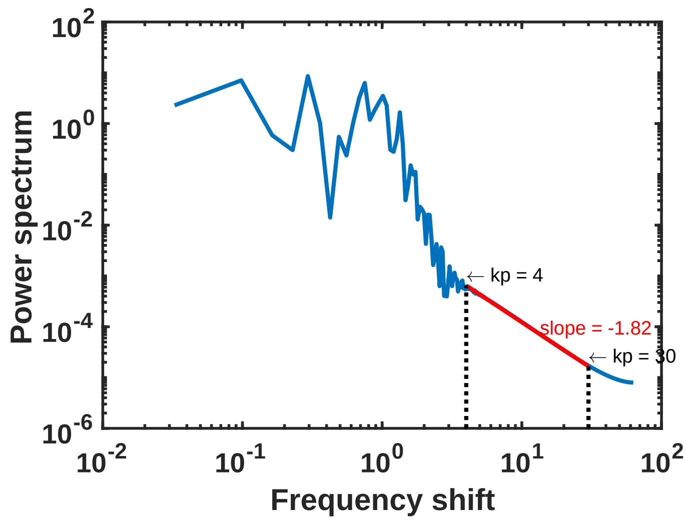
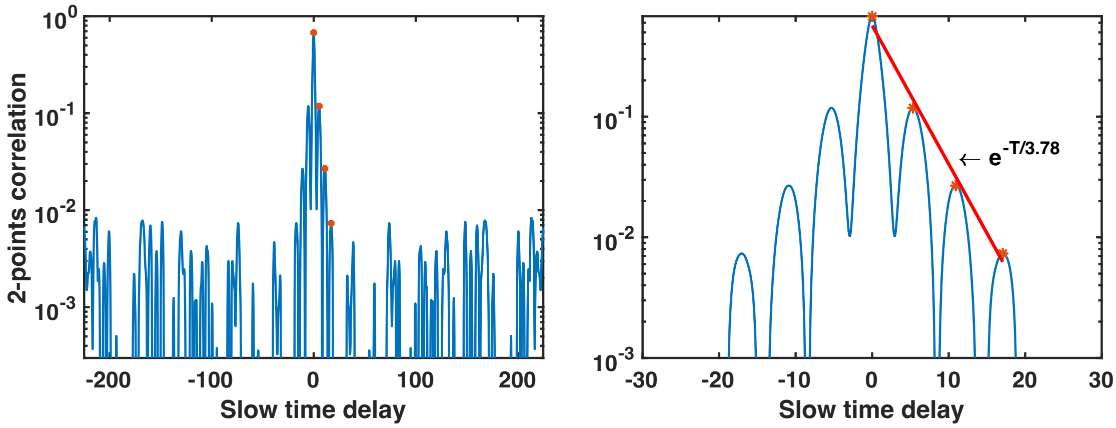


Figure 3.7: Log scale plot of the frequency spectrum of the fluctuations around the central peak of the two-points correlation function.

For the temporal correlation function, the semi-log plots in Figs. 3.8(a) and (b) provide further insights. The temporal correlation function $C(\delta t)$ is computed similarly to $C(\delta\tau)$.

The semi-log plot in Fig. 3.8(a) shows the two-points correlation function in time, highlighting an exponential decay. This decay characterizes how quickly the correlation between two points in time diminishes. The correlation length in time is taken as the inverse of the exponential decay rate, which in this case is 3.78. Fig. 3.8(b) presents a zoomed-in view, showing the detailed structure of this decay.

These analyses illustrate that phase turbulence does not originate directly from the solitons; it results from the energy emitted by these solitons during their breathing. The synchronization and large number of solitons contribute to the overall phase turbulence in the system.



(a) (b)

Figure 3.8: (a) Semi-log plot of the two-points correlation function in time, illustrating the exponential decay. (b) Zoomed-in view of (a), showing the detailed structure of the decay. The correlation length in time is the inverse of the exponential decay rate, which is 3.78.

3.4 Detection of Spatiotemporal Peaks

In the study of spatiotemporal dynamics, identifying significant peaks within the intensity field is crucial for understanding the underlying behavior and dynamics of the system. These peaks represent regions where the amplitude is notably higher compared to surrounding areas, indicating points of intense activity. Our approach to peak detection combines the identification of regional maxima with additional processing steps to ensure the retention of only the most significant peaks, while minimizing noise interference, providing a clear and detailed map of the system's dynamic behavior.

3.4.1 Enhanced Spatiotemporal Peak Detection via Prominence Analysis

To detect significant peaks in the spatiotemporal dynamic, we employed a method that combines local maxima detection with non-maximum suppression to enhance the prominence of identified peaks. The steps are as follows:

- Local Maxima Detection: We start by analyzing the intensity matrix along both the spatial and temporal dimensions to identify potential peak locations. This step highlights points where the intensity is significantly higher compared to surrounding areas, marking them as local maxima.

- Prominence Calculation: For each detected local maximum, we compute the geometric mean of their prominence values. This ensures that only points significant in both spatial and temporal dimensions are considered, effectively enhancing the prominence of the detected peaks. Prominence is a measure of a peak's relative importance compared to nearby peaks, emphasizing the most significant peaks while filtering out noise and less significant fluctuations.

- Non-Maximum Suppression: To further refine the peak detection, non-maximum suppression is applied. This process compares each peak to its neighbors within a specified neighborhood and retains only those peaks that are greater than their immediate neighbors, thereby suppressing less significant peaks and reducing noise.

- Threshold Application: A threshold is then employed to filter out peaks with low prominence, ensuring that only the most substantial peaks are retained for further analysis. This step is crucial for isolating the most active regions in the spatiotemporal field.

This method effectively identifies the most prominent features within the spatiotemporal field, providing a clear and detailed map of significant peaks. Fig. 3.9 demonstrates the detected peaks (marked in red) within the spatiotemporal map, highlighting regions of significant amplitude variations. By employing this enhanced peak detection method, we achieve a robust identification of significant peaks within the spatiotemporal dataset.

3.4.2 Inter-pulse Distance Analysis

After detecting the most significant pulses, we analyze their distribution based on their intensity. We first sort these pulses and then divide them into 10 percentile groups, from the 0^{th} to the 100^{th} percentile, creating a set of ranges that categorize the pulses into bins based on their intensity. For each

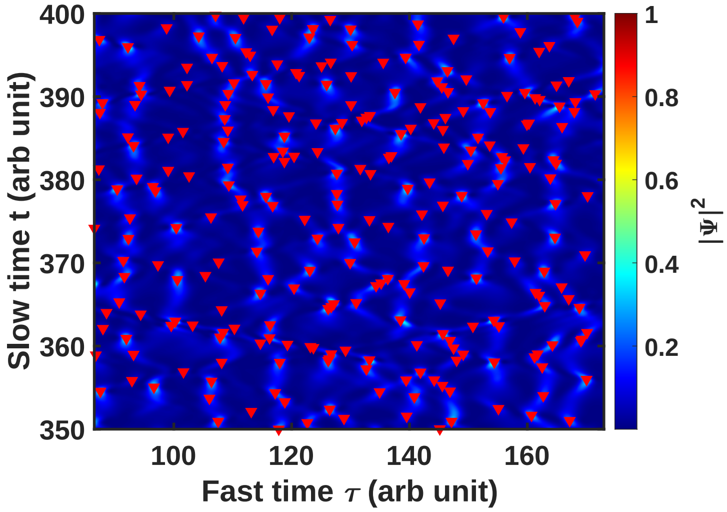


Figure 3.9: Detected significant peaks (red triangles) in the spatiotemporal map where the amplitude exhibits notable variations.

percentile group, the pulses are shifted so that their peak maximum aligns at the center, and normalized to the range $[0, 1]$, this allows for comparison between pulses of different absolute magnitudes. For each percentile group, a plot is created that shows the mean of the normalized and shifted pulses (Fig. 3.10). This represents in orange the average shape of the pulses in each percentile group. The normalized spatial two-point correlation function $C(\delta\tau)$ is also plotted in blue for comparison. By comparing the mean shapes of the pulses with the correlation function, we can assess how the detected pulses relate to the overall spatial coherence and distribution of the field. This comparison helps to identify whether the peaks are randomly distributed or if they exhibit a more structured and coherent pattern, which is crucial for understanding the dynamics and interactions within the spatiotemporal system. For each significant peak detected, we find the positions of the nearest significant peaks on either side, constrained by a predefined threshold, which is set to be 1.25 times the ratio of the mean intensity to the maximum intensity. The distances to these significant peaks from the center are then calculated. Subsequently, we compute the average distance to the left and to the right for the significant peaks within each percentile group. The results show that the average distances to the nearest significant peak to the right of a central reference peak start around 6.51 and increase to about 8.97 times the Lyapunov dimension length. For the left side, the distances start around 5.51 and increase to about 8.43 times the Lyapunov dimension length. This suggests a general trend where, as the magnitudes of the peaks increase, the average distance to the next significant peak on both sides of a central peak increases. These findings highlight the spatial distribution and coherence of significant peaks, providing insights into the structural organization of the spatiotemporal field.

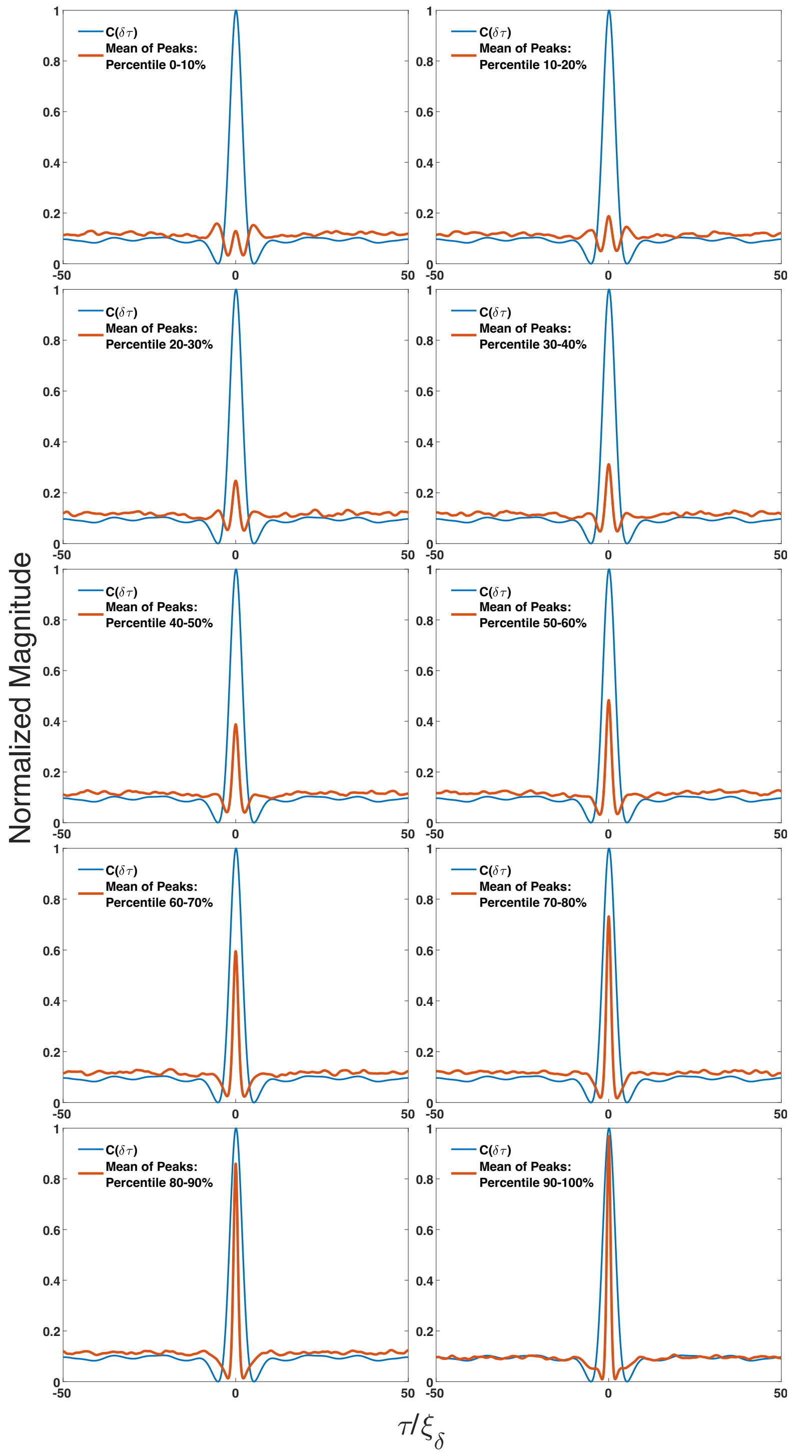


Figure 3.10: Mean of the normalized and shifted pulses for each percentile group (in orange) compared to the normalized spatial two-point correlation function (in blue).

To illustrate this, Fig. 3.11 shows an example of a central peak (marked in green) and its nearest significant peaks to the left (marked in red) and right (marked in blue). The black dashed line represents the threshold used for peak detection. This example highlights how the significant peaks are identified and their relative positions within the spatiotemporal field.

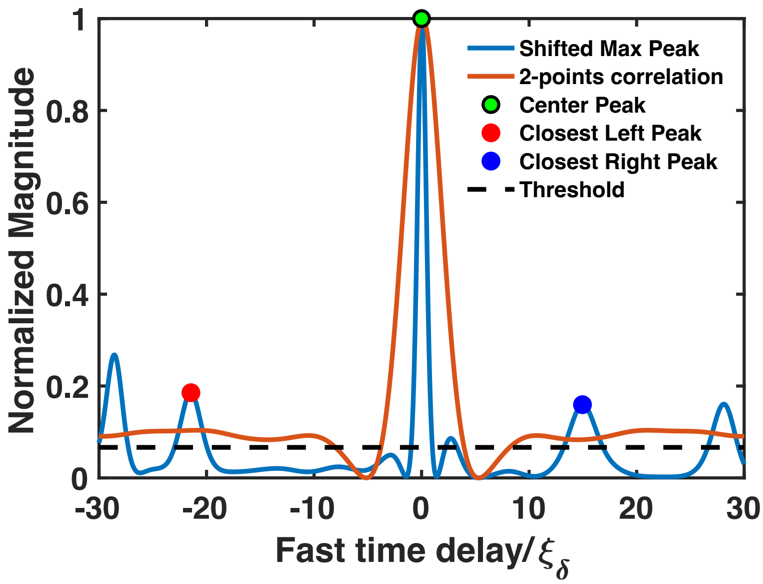


Figure 3.11: Example of a central peak and its nearest significant peaks to the left and right. The central peak is marked in green, the closest left peak in red, and the closest right peak in blue. The black dashed line represents the threshold used for peak detection.

As we will see in the next section, these significant peaks are not the only defining feature of the spatiotemporal dynamics. In addition to these high-amplitude peaks, the system also exhibits topological defects—regions where the amplitude of the field approaches zero, leading to phase singularities. The study of these defects is crucial for understanding the full extent of the turbulent behavior observed in our system.

3.5 Topological Defects and Their Role in Turbulence

In the study of spatiotemporal dynamics, defects play a critical role in understanding the underlying turbulence phenomena. Defects are points in the system where the amplitude of the field nearly vanishes, leading to simultaneous crossings of zero by both the real and imaginary parts of the field. This results in a phase singularity at these points. To identify these defects, we search for local minima in the amplitude of the spatiotemporal field, which indicates regions where the phase is not well-defined due to the vanishing amplitude. Once potential defects are identified, it is crucial to verify their nature as true spatiotemporal defects. This is done by examining the contour around the identified defect. Specifically, we calculate the contour integral of the phase around the defect. If this integral equals 2π or a multiple thereof, it confirms the presence of a phase singularity, thereby validating the defect as a true topological defect. Fig. 3.13 presents a zoomed-in view of one such defect, focusing on the phase dynamics around the defect. Specifically, we calculate the contour integral of the phase around this defect. If this integral equals 2π or a multiple thereof, it confirms the presence of a phase singularity, thereby validating the defect as a true topological defect.

The presence of defects in a turbulent system indicates regions of phase turbulence. Phase turbulence is characterized by continuous changes in the phase of the field, leading to complex spatiotemporal patterns. The alternation between phase turbulence and defect-mediated turbulence suggests a rich, multi-faceted dynamical behavior. This alternation can lead to the emergence of extreme events, which are rare, high-amplitude excursions in the system's dynamics. Understanding these defects and their role in the dynamics provides valuable insights into the mechanisms driving spatiotemporal chaos and turbulence in such systems.

In our study, we started with a soliton and gradually increased the pump parameter S until $S = 0.4$. This transition led to the onset of spatiotemporal chaos, marked by the emergence of numerous defects. When S was subsequently decreased, the spatiotemporal chaos persisted even at lower values of S where initially there was no chaos, demonstrating a hysteresis effect. This indicates that the system maintains its chaotic state despite the reduction in S to levels that originally supported solitons. Such behavior has been observed in similar systems, where a bifurcation diagram reveals the coexistence of spatiotemporal chaos, pulsating localized structures, and homogeneous steady states [102].

Fig. 3.12(a) shows a spatiotemporal map of the system, with defects highlighted as red squares. This visualization helps to identify the locations of phase singularities within the dynamic field, illustrating the distribution and density of defects over time. The defects' presence and distribution are indicative of the underlying turbulent nature of the system, providing a clear visual representation of how turbulence manifests through phase singularities and amplitude variations. Fig. 3.12(b) depicts the number of defects as a function of the pump parameter S, showing a hysteresis loop. The circles represent the incremented S values, and the triangles represent the decremented S values, highlighting the difference in defect numbers during the increasing and decreasing phases of S.

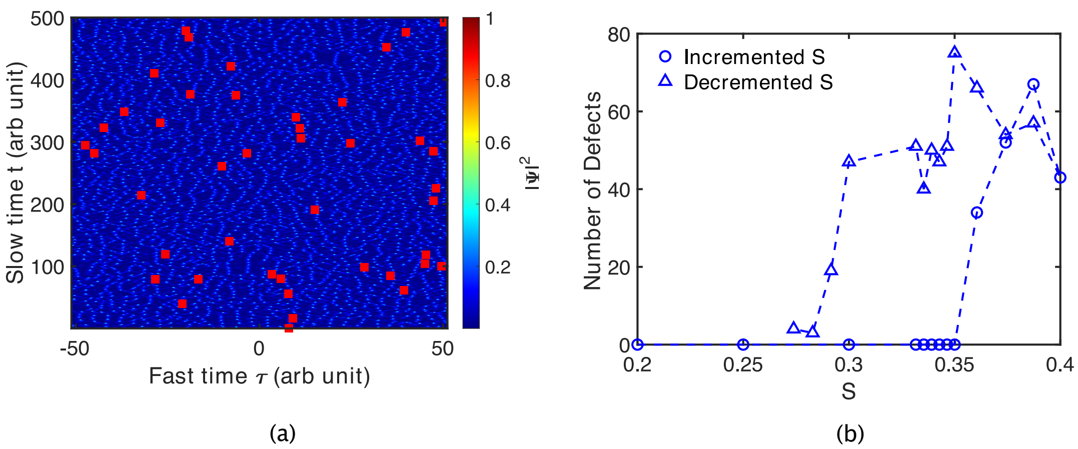


Figure 3.12: (a) Spatiotemporal map of the system showing defects (red squares) where the amplitude of the field nearly vanishes. (b) Number of defects per fast and slow time units as a function of the pump parameter S, illustrating the hysteresis loop with circles for incremented S and triangles for decremented S.

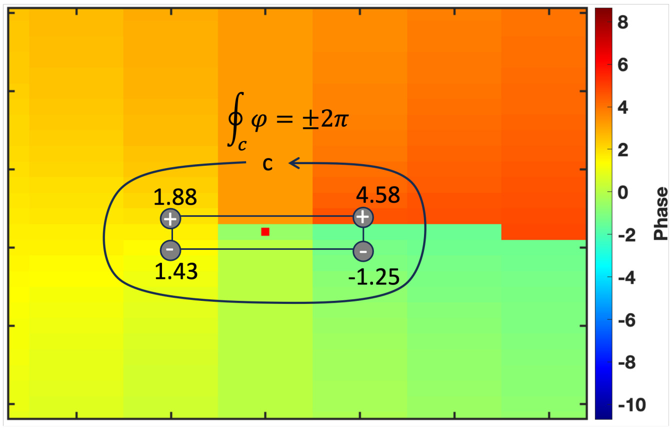


Figure 3.13: Zoomed-in view of a defect showing the phase dynamics around it. The contour integral of the phase calculated around the defect confirms the presence of a phase singularity, with the integral equaling 2π, validating the defect as a true topological defect.

3.6 Structure Functions and Multiscale Dynamics

The study of the higher-order correlations, commonly referred to as structure functions is essential for unraveling the complex, multi-scale nature of turbulence. Structure functions provide a detailed statistical characterization of the turbulent field, revealing information about higher-order correlations and the presence of characteristic scales. By analyzing the scaling behavior of these functions, we can gain insights into self-similarity and intermittency, which are key features of turbulent systems [98].

To this end, we examined the structure functions, which are mathematically defined for order p as follows:

$$S_p^{(d)}(\tau) = \langle \|\delta_\tau^{(d)} I(t)\|^p \rangle_l, \tag{3.7}$$

$$\delta_\tau^{(l)} I(t) = I(t+\tau) - I(t), \tag{3.8}$$

$$S_p^{(d)}(\tau) \sim \tau^{\zeta_p}. \tag{3.9}$$

In these equations, $S_p^{(d)}(\tau)$ represents the structure function of order p, a statistical measure quantifying the differences in the intensity field $I(t)$ over a time lag τ. The angle brackets $\langle \cdot \rangle_l$ denote an average over the spatial domain. The incremental difference in intensity over a time lag τ is given by $\delta_\tau^{(l)} I(t)$. The power-law scaling relationship characterized by the exponent ζ_p is indicative of the underlying self-similarity or intermittency within the turbulent field. According to the Kolmogorov theory of turbulence, for fully developed turbulence, the p^{th} order structure function scales with the separation distance r in the inertial subrange as $r^{\frac{p}{3}}$. This theoretical framework provides a foundational understanding of turbulence; however, empirical observations frequently exhibit deviations from this simple scaling law, particularly for higher-order structure functions. Such deviations are typically attributed to the phenomenon of intermittency, characterized by irregular, sporadic occurrences of intense small-scale fluctuations within the flow [98].

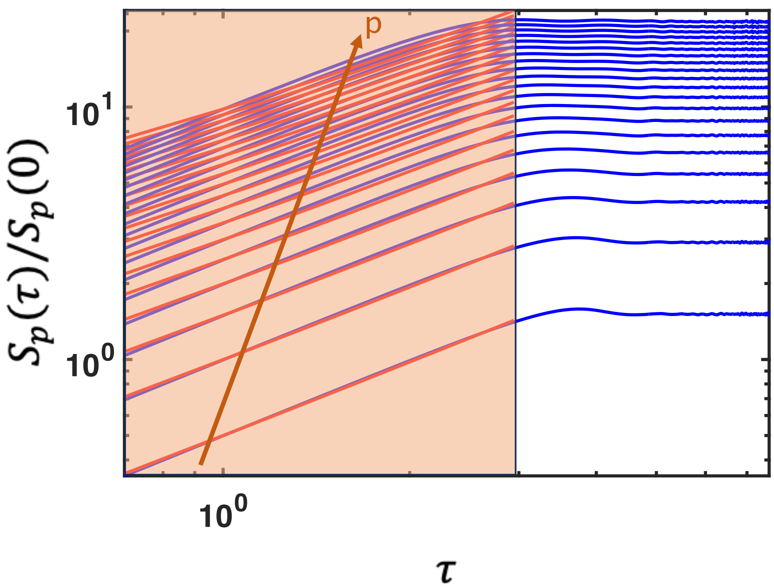
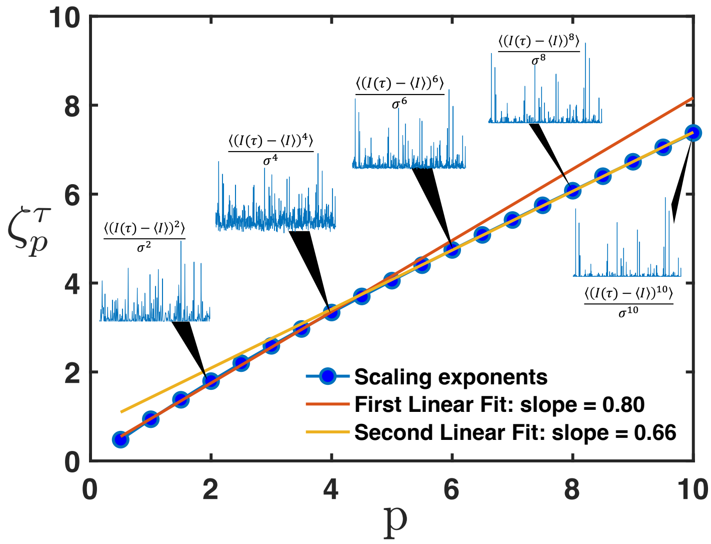
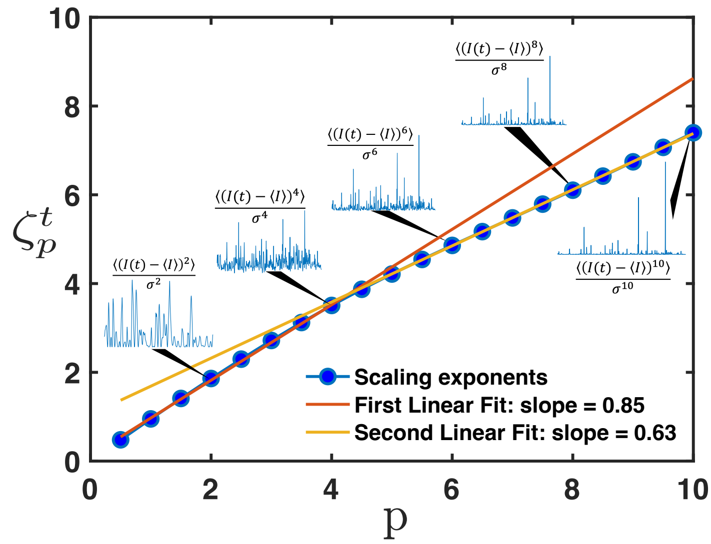


Figure 3.14: Spatial structure functions $S_p(\tau)$ calculated for various orders p.

In this work, we calculate the structure functions and scaling exponents of the field intensity. We observe that both spatially and temporally a close to 0.8 law is well followed until order $p = 4.5$, from this critical value the curve bends away. Therefore, the dynamic exhibits intermittent behavior, demonstrating deviations from the expected scaling law at higher orders, reflecting the complex nature of turbulent systems. Fig. 3.14 shows the spatial structure functions, while Figs. 3.15a and 3.15b depict the spatial and temporal scaling exponents, respectively, showing the intermittent behavior.



(a)



(b)

Figure 3.15: Scaling exponents (a) ζ_p^τ and (b) ζ_p^t of the spatial and temporal structure functions, respectively, demonstrating intermittent behavior.

To better illustrate this dynamical behavior, we also present profiles of the moments m_p for different exponents p, as seen in the insets of Fig. 3.15a. The moments are calculated as: $m_p \equiv \langle (I(\tau) - \langle I \rangle)^p \rangle / \sigma^p$, where $I(\tau)$ is the total intensity at position τ, $\langle I \rangle$ is the average intensity, and σ is the standard deviation. The insets clearly show the loss of self-similarity at $p = 4.5$, followed by intermittent behavior. A similar analysis was conducted for temporal fluctuations, as shown in the insets of Fig. 3.15b, where the moments $m_p^{'}$ were calculated for different exponents p, where $m_p^{'} \equiv \langle (I(t) - \langle I \rangle)^p \rangle / \sigma^p$, with $I(t)$ is the total intensity at time t and $\langle I \rangle$ is the average intensity.

The analysis of structure functions provides deep insights into the multi-scale nature of turbulence and helps clarify the transition between self-similar and intermittent dynamics. This understanding is vital for advancing our knowledge of turbulent systems.

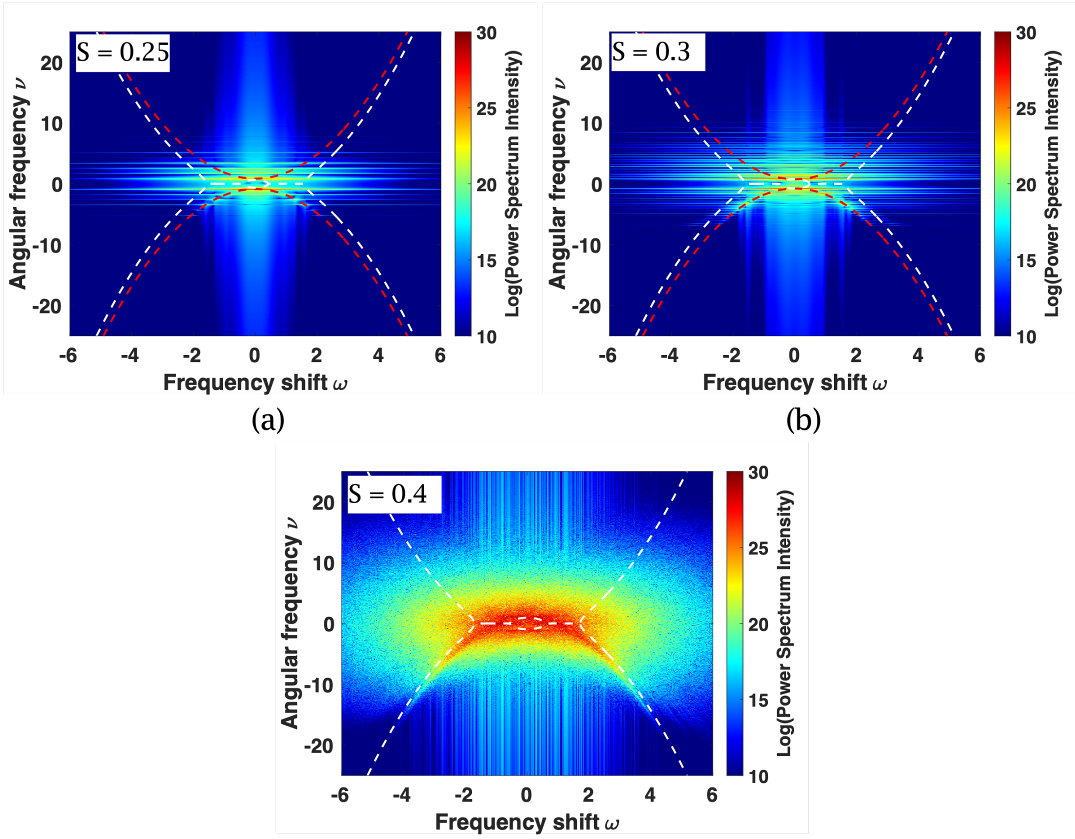
3.7 Link to Dispersion Relation

We now revisit the dispersion relation of the LLE, which characterizes the perturbation of the Homogeneous Steady State (HSS) solution, given by the following form:

$$\lambda_{\pm} = -\alpha \pm \sqrt{I_s^2 - (\Delta - 2I_s + \omega^2)^2} \tag{3.10}$$

In this expression, $\lambda_{\pm}$ are the eigenvalues derived from the linear stability analysis, α represents the cavity loss parameter, I_s is the intracavity field intensity, Δ denotes the detuning, as discussed in Section 1.3.2. Figs. 3.16(a)-(c) illustrate the logarithmic power spectrum of the Fourier transform applied in both the slow and fast time directions for various values of the pump parameter S (0.25, 0.3, and 0.4, respectively). The overlaid red and white curves represent the imaginary components of the eigenvalues from the dispersion relation. The white curves correspond to the imaginary parts of $\lambda_{\pm}$ for the value of I_s that lies on the upper branch of the bistability hysteresis curve, while the red curves correspond to the lower branch. The horizontal axis represents the frequency shift ω, while the vertical axis corresponds to the angular frequency ν (imaginary part of λ).

From Figs. 3.16, it is evident that the soliton spectrum dominates in intensity,



(c)

Figure 3.16: Dispersion relation of the Lugiato-Lefever Equation (LLE) for different pump parameters S. The logarithmic power spectrum of the Fourier transform in both slow and fast time directions is shown for (a) $S = 0.25$, (b) $S = 0.3$, and (c) $S = 0.4$. The overlaid curves represent the imaginary parts of the eigenvalues from the dispersion relation.

with the trace of the dispersion relation clearly visible, which is linked to the presence of stable linear waves propagating along the HSS solution. Each time a soliton dissipates its energy, the emitted waves must follow the dispersion relation corresponding to the perturbation of the HSS; any other wave that does not satisfy the dispersion relation cannot propagate. This explains why the spectra in these figures remain well-defined, rather than appearing noisy.

However, in Fig. 3.16(c), corresponding to $S = 0.4$, the spectrum broadens, reflecting the emission of a high number of solitons. These solitons are not necessarily synchronized in time, and even when they coexist, they do not share the same phase, leading to a lack of synchronized behavior. Despite this, far from the soliton region, the waves propagate on the HSS solution.

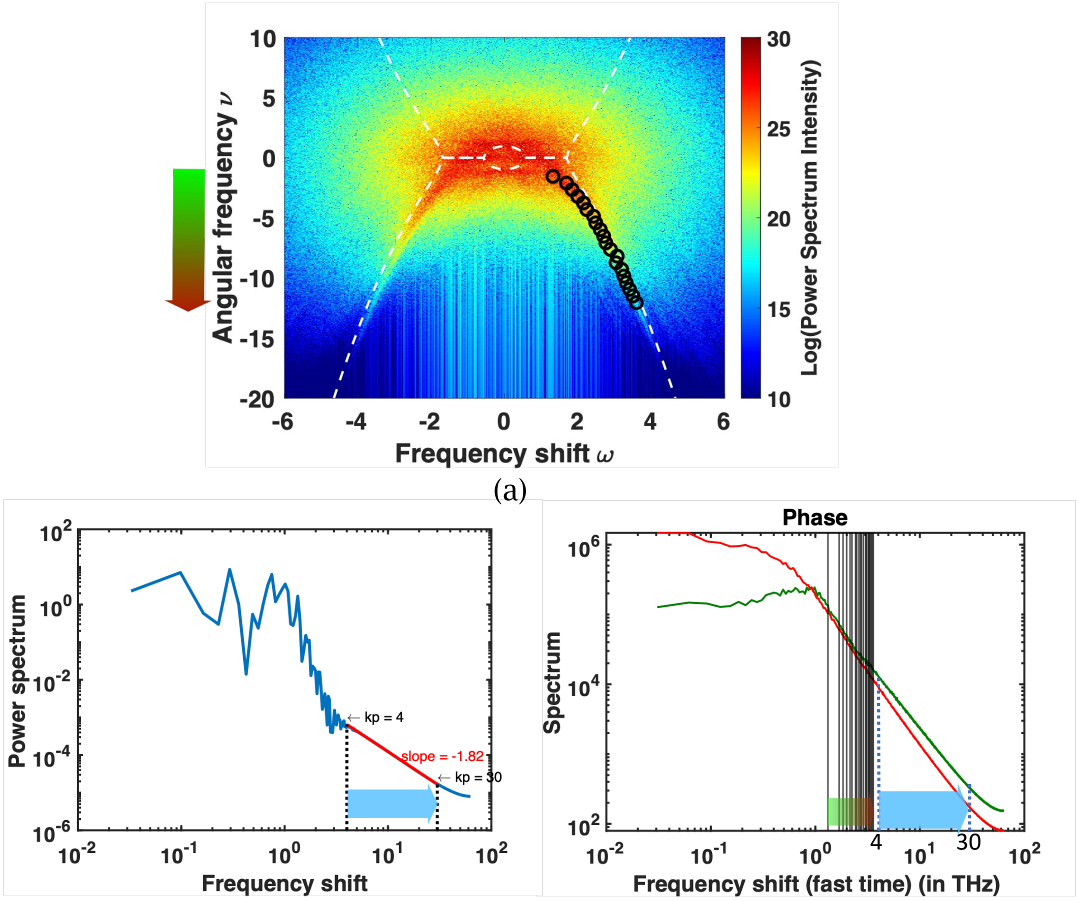
The solitons themselves are unstable. When they appear, they undergo a Hopf instability, leading to oscillations that eventually destroy the solitons, resulting in the emission of waves. Since the solitons appear and disappear, the formation of one is sufficiently distant from the previous one such that, when it dissipates, its energy propagates. This behavior was observed when we computed the inter-pulse distance in Section 3.4.2.

3.7.1 Detailed Analysis of Dispersion Relations and Spectral Behavior

The dynamics in this system represent a mixture of breathing cavity solitons and linear waves propagating along the HSS solution. However, to understand which domain contributes to the energy cascade of the phase, we highlight in Fig. 3.16c the corresponding frequency shift values at specific angular frequencies where the power spectrum intensity is maximized, as shown by the black circles in Fig. 3.17a. The gradient arrow on the left indicates increasing angular frequencies (ν), helping to visualize where these frequencies contribute to the energy cascade, as seen in Fig. 3.17c. It becomes evident that these linear waves are responsible for initiating the energy cascade.

Figure 3.17b corresponds to the same spectrum as in Fig. 3.7 from Section 3.3, which illustrates the fluctuations in intensity. The spectrum, obtained from the Fourier transform of these fluctuations, again reveals a power-law decay with an exponent close to -1.82, suggesting self-similar dynamics or phase turbulence within the system. Once we move beyond the correlation domain of the localized structures, we recover the energy cascade from the fluctuations. The solitons, however, prevent these slow fluctuations from contributing to an energy cascade.

The region of frequency shifts displaying power-law decay is clearly identified and can be located within the spatial spectrum of the phase in Fig. 3.17c, alongside the region corresponding to linear waves. (note that this region can extend more if we take into account also the high-frequency shifts).



(a) (b) (c)

Figure 3.17: (a) Dispersion relation of the LLE for $S = 0.4$. Black circles highlight the corresponding frequency shift values for specific angular frequencies. (b) Power spectrum in logarithmic scale as a function of the frequency shift of the intensity fluctuations showing a universal exponent close to -2. (c) The spatial spectrum of the phase, showing the frequency shifts corresponding to both the linear waves from (a) and those corresponding to the intensity fluctuations from (b).

3.8 Conclusion

In this chapter, we delved into the complex dynamics of a system operating in the bistable regime, concentrating on the transition from soliton-based structures to spatiotemporal chaos and fully developed turbulence. As we increased the pump parameter, the system evolved from a soliton-based regime to a spatiotemporal chaotic one characterized by phase turbulence. Through the analysis of phase and amplitude power spectra, we identified distinct scaling laws. The phase dynamics, in particular, exhibited a power-law behavior indicative of phase turbulence, while the amplitude decayed more steeply, suggesting differing underlying mechanisms in the turbulent regime.

To understand the difference between the phase and amplitude dynamics, we conducted a detailed examination of the dispersion relation of the Lugiato-Lefever Equation (LLE), we identified the critical role that stable linear waves, propagating along the Homogeneous Steady State (HSS) solution, play in initiating the energy cascade of the phase spectrum. Further analysis using two-point correlation functions showed that, outside the correlation domain of localized structures, the fluctuations spectrum follows a power-law decay.

Furthermore, the analysis using structure functions allowed us to explore higher-order correlations, revealing that the system follows a scaling law up to a critical power. Beyond this point, the system diverges from the observed scaling, indicating that it becomes intermittent. This transition signifies that at larger scales, the dynamics become irregular, marking the onset of intermittency. Another key finding of this chapter is the emergence of topological defects—phase singularities, serve as markers of turbulence and contribute to the formation of extreme events.

Overall, this chapter has provided a comprehensive analysis of the transition from soliton dynamics to fully developed turbulence, offering new insights into the mechanisms governing spatiotemporal chaos in optical systems.

Chapter 4

Advanced Analysis of Wind Dynamics Using Transfer Entropy

Outline of the current chapter

In this chapter, we delve into the advanced analysis of wind dynamics using the concept of Transfer Entropy (TE) to uncover the directional flow of information between different meteorological stations. The chapter begins with an introduction to the importance of understanding wind behavior for applications such as weather forecasting and renewable energy optimization. We then describe the data collection and preprocessing steps, including handling missing data and time gaps. The main focus is on analyzing wind speed and direction using both univariate and multivariate TE methods. Univariate TE examines the information transfer for wind speed and direction separately, while multivariate TE considers the combined influence of wind speed and direction components. Through detailed examples and comparisons, we illustrate how TE can reveal complex interdependencies and directional influences among stations. The chapter concludes by highlighting the significant findings and the robustness of TE as a tool for enhancing our understanding of wind dynamics and improving predictive models.

4.1 Introduction

Wind dynamics play a crucial role in various aspects of atmospheric sciences, impacting weather patterns, climate models, and environmental monitoring. Accurate understanding and prediction of wind behavior are essential for applications ranging from weather forecasting to renewable energy optimization. In this chapter, we focus on analyzing wind data from multiple meteorological stations in Sydney, Australia, using advanced information theory methods to understand the directional flow of information. Wind data analysis is a fundamental aspect of atmospheric sciences, as wind patterns influence countless environmental and climatic processes. The study of wind dynamics involves understanding the movement of air masses across various spatial and temporal scales, which can significantly impact weather prediction, climate modeling, and renewable energy resource management. Wind data is typically collected from meteorological stations that measure wind speed and direction. These measurements are crucial for developing accurate weather models. Additionally, understanding wind behavior is essential for designing and optimizing wind farms, which are a key component of renewable energy strategies.

Wind data analysis faces several challenges, such as dealing with missing or incomplete data [189], accounting for complex terrain effects [190], and capturing the inherently chaotic nature of atmospheric processes. As a result, researchers have developed various statistical and computational methods to enhance the analysis and interpretation of wind data. One of the primary tools in wind data analysis is time series analysis, involving the examination of historical wind data to detect periodicities and trends. Techniques such as autoregressive integrated moving average (ARIMA) models have been widely used to forecast wind speed and direction based on past observations [191]. While these methods provide valuable insights, they often assume linear relationships and may not adequately capture the complex, nonlinear interactions present in atmospheric systems. Numerous studies have explored wind speed and direction analysis, employing various methodologies to understand wind dynamics and improve forecasting models. Early research primarily focused on statistical techniques such as correlation analysis and regression models to identify patterns and trends in wind data [192, 193].

Bilgili et al. (2007) [194] used correlation analysis based on the evolution of the sample cross-correlation function (SCCF) from [195] to select reference and target stations for wind speed prediction using Artificial Neural Networks (ANN). While this method provides valuable insights, it assumes linear relationships and does not account for the directional flow of information. Consequently, some stations that may appear uncorrelated could still exchange significant information. To address the limitations of traditional statistical methods, we propose using Transfer Entropy (TE) to analyze wind data. TE, introduced by Schreiber in 2000 [155], is a non-parametric measure that quantifies the directed transfer of information between two time series. Unlike traditional methods, TE can detect nonlinear dependencies and delayed interactions, making it particularly suited for analyzing the intricate dynamics of wind patterns.

Several studies have successfully applied TE to wind data analysis. For example, Ragwitz and Kantz (2002) demonstrated the utility of TE in detecting causal relationships in chaotic systems, including atmospheric dynamics [196]. More recently, other studies have used TE to study the directional information flow in wind speed data across multiple meteorological stations, highlighting the method's ability to uncover complex interdependencies. The integration of information theory-based methods with traditional statistical techniques represents a promising direction for wind data analysis. By leveraging the strengths of both approaches, researchers can develop more accurate and robust models for predicting wind behavior, ultimately contributing to improved weather forecasting and renewable energy management.

Our primary objective in this chapter is to explore the directional information flow within wind data, which encompasses both wind speed and direction. By applying TE [155], we aim to uncover how information propagates between different meteorological stations. This analysis is pivotal for enhancing our understanding of wind dynamics and improving forecasting models. Transfer Entropy, discussed in Section 1.6.2, is a non-parametric, model-free measure that captures the directed transfer of information between two systems. Unlike traditional methods such as Granger causality [162], which often assume linear relationships, TE is capable of detecting nonlinear interactions, making it particularly suited for complex systems like atmospheric dynamics. By examining both univariate TE (focusing on wind speed and direction separately) and multivariate TE (considering the combined effect of wind speed and direction), we can obtain a comprehensive view of information flow.

Overall, the integration of information theory-based methods with traditional statistical techniques represents a promising direction for wind data analysis. By leveraging the strengths of both approaches, researchers can develop more accurate and robust models for predicting wind behavior, ultimately contributing to improved weather forecasting and renewable energy management.

4.2 Data Collection and Preprocessing

4.2.1 Meteorological Stations and Data Description

The data for this study was collected from a network of 57 meteorological stations distributed across Sydney, Australia, as shown on the map in Fig. 4.1. These stations are equipped with instruments that measure various atmospheric parameters, including wind speed and direction. Wind speed is the average wind speed in meters per second (m/s), and wind direction is the average wind direction over the measurement minute in degrees azimuth. The data was recorded at one-minute intervals throughout the year 2021, providing a high-resolution dataset suitable for detailed analysis. The geographic distribution of these stations allows for a comprehensive analysis of wind patterns across different areas, capturing local variations influenced by geographic features such as coastlines, urban areas, and topographical elevations.

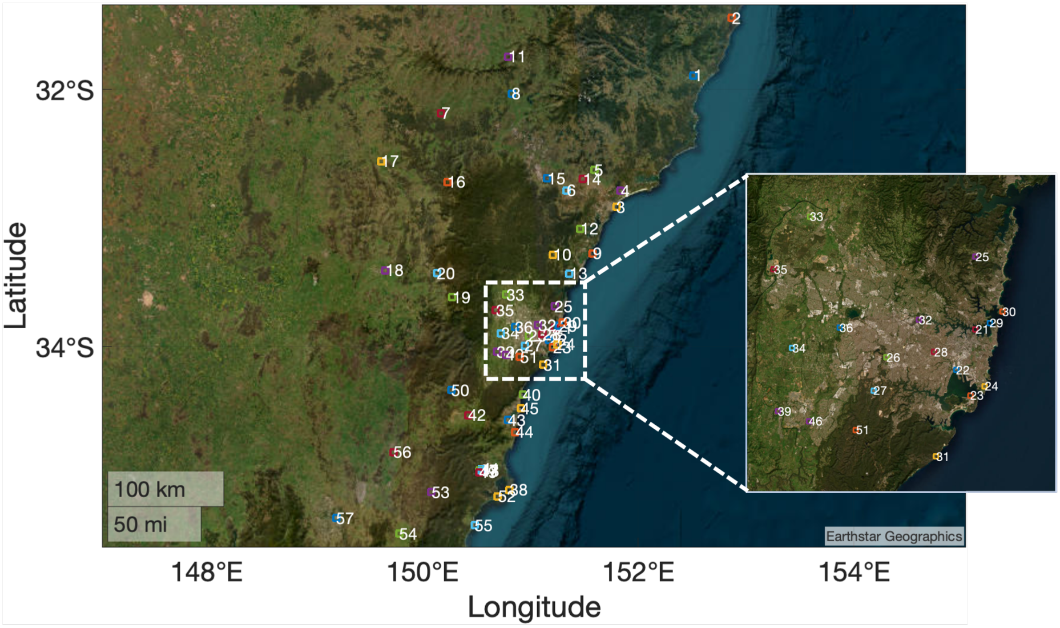


Figure 4.1: Map showing the distribution of the 57 meteorological stations across Sydney, Australia. The inset provides a zoomed-in view of the area with a higher density of stations, for better clarity.

4.2.2 Handling Missing Data and Time Gaps

To ensure the consistency and reliability of the dataset, we performed an initial cleaning process. The first step involved removing the data for January 2021 because this month had missing entries for some stations. This decision was made to maintain uniformity in the number of stations across the entire dataset.

Despite the comprehensive coverage, the dataset exhibited some irregularities, including time gaps and NaN (Not a Number) values. These gaps and NaNs indicate periods when the recording instruments were not operational or data was not recorded accurately. Although there are methods to handle these NaN values, they are not efficient for large gaps. For our analysis, it is crucial to work with continuous time series data, as TE calculations require uninterrupted sequences to provide meaningful results. Simply replacing NaN values with zeros was not an option because a zero value in wind speed data would inaccurately indicate no wind, which could be a valid observation rather than a data gap. Instead, we adopted a more sophisticated approach to handle these issues. We identified all time gaps and instances of NaN values within the dataset by checking for missing timestamps and scanning for NaNs in the wind speed and direction measurements. All identified time gaps were filled with NaN values to explicitly mark periods with missing data, ensuring that any gaps would not be overlooked during the analysis.

Next, we searched for the longest sequences of continuous data, where no NaN values were present. These sequences were extracted for subsequent analysis to ensure that TE calculations would be based on reliable and uninterrupted data. This involved segmenting the longest continuous data segments for each station and storing them separately.

This preprocessing step is critical as it directly impacts the accuracy and reliability of the TE analysis. By focusing on the longest continuous sequences, we aimed to maximize the integrity and utility of the dataset for studying the directional information flow in wind data. This preprocessing ensures that the subsequent TE calculations are performed on uninterrupted data segments, thereby enhancing the robustness of our analysis and findings.

4.3 Transfer Entropy Analysis

4.3.1 Univariate TE for Wind Speed and Direction

Transfer Entropy is a powerful tool for analyzing the directional transfer of information between two time series. In the context of wind data, univariate

TE can be applied separately to wind speed and wind direction to uncover how information flows between different meteorological stations. Univariate TE measures the amount of information transferred from one time series to another, considering each variable independently. For wind speed, we compute the TE from the wind speed at one station to the wind speed at another station. Likewise, for wind direction, we compute the TE from the wind direction at one station to the wind direction at another station. This helps identify directional dependencies and potential causal relationships between stations.

Univariate TE has been widely used in various fields to study directional dependencies and causality. In atmospheric sciences, it helps in understanding how wind patterns propagate between different locations, which is crucial for weather forecasting and climate modeling.

The analysis involves computing TE for various time lags to identify the optimal lag at which the information transfer is most significant. This is done by systematically shifting the time series of one station and calculating the TE with the time series of another station at each shift. The time lag with the highest TE value indicates the delay at which the information transfer is maximized. For two time series X and Y, the computation of TE in this context from X to Y at a lag τ can be mathematically represented as follows:

$$TE_{X \to Y}(\tau) = \sum p(y_{t+\tau}, y_t, x_t) \log \frac{p(y_{t+\tau}|y_t, x_t)}{p(y_{t+\tau}|y_t)} \tag{4.1}$$

where $y_{t+\tau}$ is the future value of Y, y_t is the current value of Y, and x_t is the current value of X. This measure captures the reduction in uncertainty of the future state of Y given the past state of X and Y, compared to the uncertainty of Y given only its past state. A higher TE value indicates a stronger influence of X on Y. The computation of TE involves estimating the joint and conditional probability distributions of the time series data. By applying univariate TE, we aim to uncover the complex interactions and directional dependencies in wind data, providing valuable insights for improving wind forecasting models and understanding atmospheric dynamics.

4.3.2 Multivariate TE for Wind Direction Components

To capture the combined influence of wind speed and direction, we transform the wind data into its x and y components. This transformation allows for a more detailed analysis of the directional information flow between stations by considering the vector components of wind. The Wind Speed (WS) in meters per second (m/s) and Wind Direction (WD) in degrees azimuth[1] are used to compute the x and y components of the wind vector:

$$\begin{cases} x = \mathrm{WS} \times \cos(\theta) \\ y = \mathrm{WS} \times \sin(\theta) \end{cases} \tag{4.2}$$

where θ is the Wind Direction (WD) in radians.

By analyzing these components, multivariate TE provides a comprehensive view of how wind information flows between stations. In the multivariate TE analysis, X and Y are now two-dimensional vectors instead of one-dimensional. This means that instead of considering only the wind speed or direction at a station, we now consider both the x and y components of the wind vector at each station. Multivariate TE extends the univariate approach by considering multiple variables simultaneously. This approach captures more complex interactions and offers a richer analysis of information flow by considering the combined influence of wind speed and direction. In multivariate TE analysis, the joint influence of multiple time series is examined to understand how they collectively affect another time series. This is particularly useful in wind data analysis, where both wind speed and direction play a crucial role in determining wind patterns. By incorporating the x and y components of the wind vector, multivariate TE can capture the combined influence of these variables on the information transfer between stations. Multivariate TE is particularly powerful for analyzing complex systems with multiple interacting variables. The methodology employed here for calculating Multivariate TE was rigorously tested and validated in Appendix B using coupled Rössler systems, ensuring its robustness and reliability for complex systems like wind dynamics. In the context of wind

[1] In degrees azimuth, 0° corresponds to North, 90° to East, 180° to South, 270° to West, and the degrees increase clockwise.

data, it allows us to understand how the combined effects of wind speed and direction at different locations influence wind patterns across the network of meteorological stations.

Overall, both univariate and multivariate TE analyses provide valuable insights into the directional information flow in wind data, helping to improve our understanding of wind dynamics and enhancing the accuracy of wind forecasting models. The integration of these methods offers a comprehensive approach to studying wind patterns and their interactions across different spatial and temporal scales.

4.4 Wind Speed

4.4.1 Filter and smooth the Data

To filter the wind speed data, we used a combination of the Savitzky-Golay Smoothing Filter [197] and a moving mean filter with a window size of 5 min. The Savitzky-Golay Smoothing Filter applies a polynomial regression to a moving window of data points, effectively smoothing the data while preserving important features such as peaks and troughs. This method is particularly useful for maintaining the integrity of the wind speed's inherent variability while reducing the noise. Additionally, the moving mean filter further reduces noise by averaging the wind speed measurements over the specified window. This step is crucial for mitigating short-term fluctuations and providing a clearer representation of the underlying wind speed trends. By combining these two filtering techniques, we achieve a balance between noise reduction and the preservation of significant wind speed features, ensuring more accurate and reliable analysis.

Let's take an example of two stations as shown in Fig. 4.2. Station 24 is on the coastline, and Station 34 is further inland. Fig. 4.3 illustrates the wind speed data before and after applying the filtering techniques. Fig. (4.3a) shows the raw wind speed data recorded between Stations 34 and 24 for a specific period. Fig. (4.3b) presents the filtered wind speed data, demonstrating the effectiveness of the Savitzky-Golay and moving mean filters in smoothing the data. Overall, we can see that the wind speed at Station 24 is consistently higher than at Station

34. This difference is primarily due to the locations of the two stations. Station 24, being on the coastline, is subject to stronger and more consistent winds due to the influence of the sea breeze and reduced surface friction. In contrast, Station 34, located further inland, experiences more variable and generally weaker wind speeds due to increased surface friction and complex topographical features.

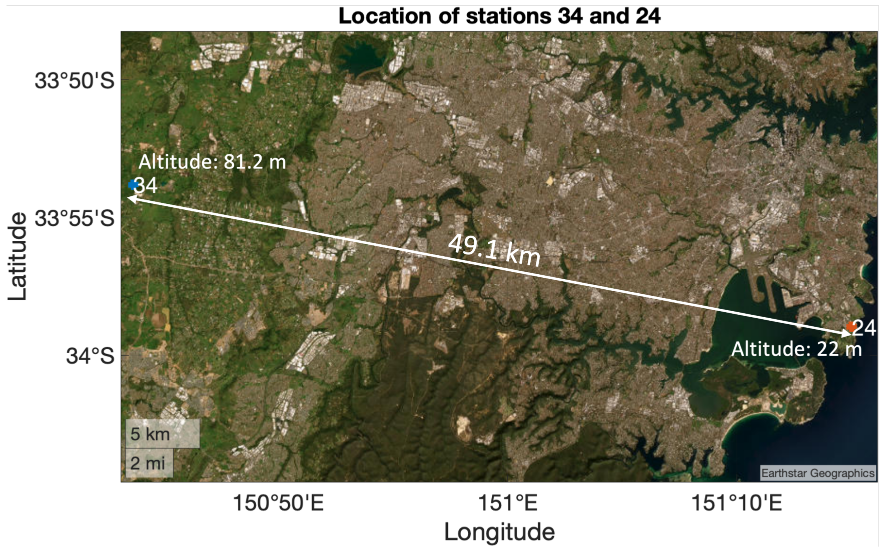
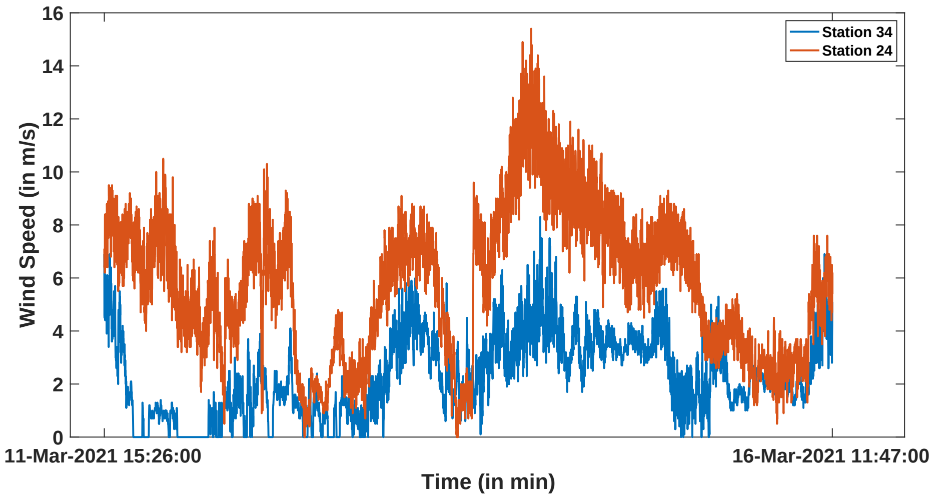


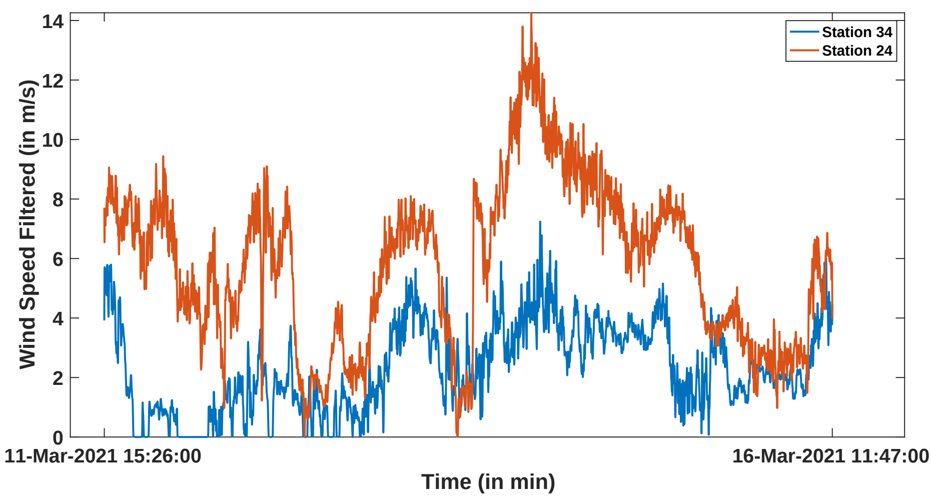
Figure 4.2: Map showing the locations of Stations 34 and 24. Station 24 is on the coastline at an elevation of 22 meters, while Station 34 is further inland at an elevation of 81.2 meters. The distance between the two stations is approximately 49.1 kilometers.

4.4.2 Transfer Entropy of wind speed

By computing the Transfer Entropy of the wind speed between the two stations 34 and 24, shown in Fig. 4.4, we can compare the TE of the raw data and the filtered data of the wind speed for different lags. Overall, we observe that the filtering did not significantly affect the overall shape of the TE values, though it did affect the values themselves. This is because the primary purpose of the filtering process is to reduce noise while preserving the fundamental relationships and information flow between the stations. The TE values reflect these relationships, and the shape of the TE curve indicates how information transfer varies with lag. The TE plots illustrate that the wind speed at Station



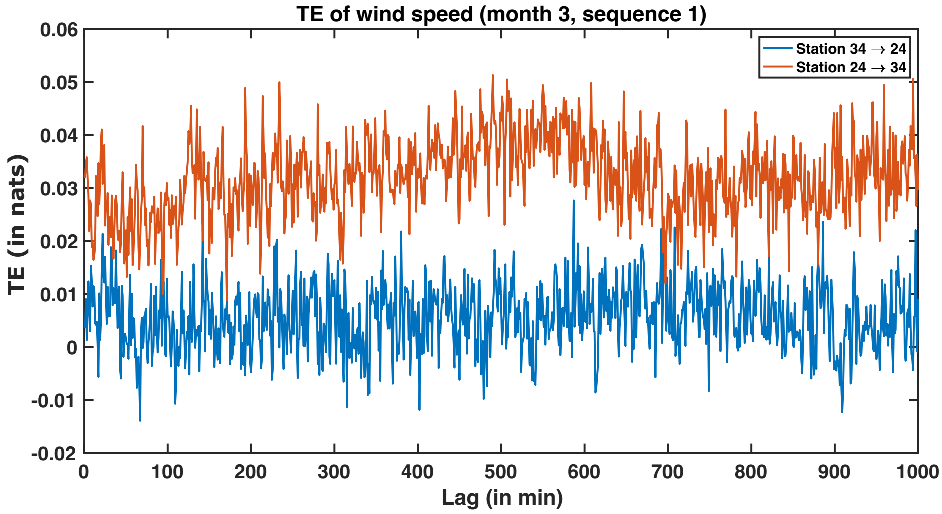
(a)



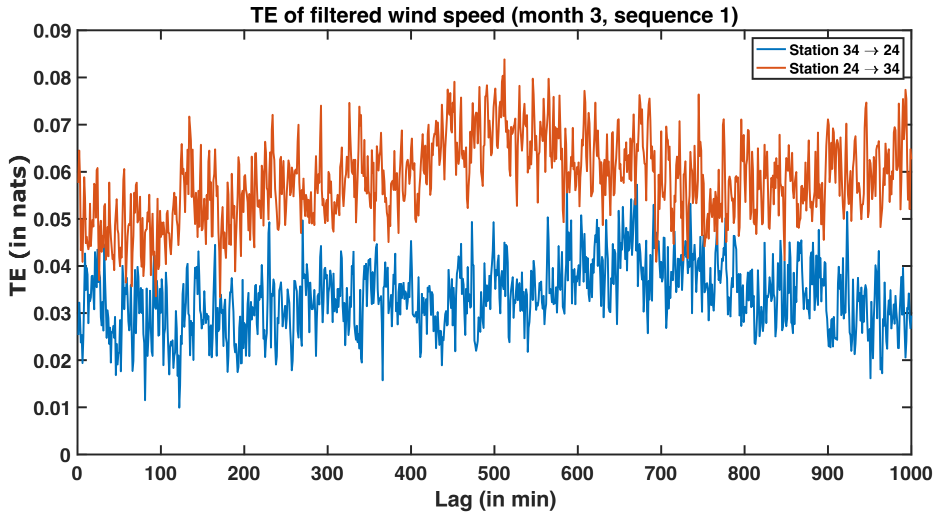
(b)

Figure 4.3: Wind speed data comparison: (a) Raw wind speed data, and (b) Filtered wind speed data using the Savitzky-Golay Smoothing Filter and moving mean filter recorded between stations 34 and 24 for a specific period.

24 (coastal) exerts a stronger influence on the wind speed at Station 34 (inland), as indicated by higher TE values from Station 24 to Station 34. This is consistent with our understanding of wind dynamics, where coastal stations typically



(a)



(b)

Figure 4.4: TE comparison: (a) Raw wind speed data, and (b) Filtered wind speed data using the Savitzky-Golay Smoothing Filter and moving mean filter recorded between stations 34 and 24 for a specific period.

experience more consistent and stronger winds, which can influence the wind patterns further inland. In contrast, the influence from Station 34 to Station 24 is weaker, reflected in lower TE values. This is due to the more variable and often weaker wind conditions inland. The consistent behavior observed in both raw and filtered TE plots suggests that the filtering process preserves the essential information flow characteristics between the stations, confirming the robustness of the filtering techniques used. This analysis highlights the importance of considering both the geographical context and appropriate data filtering techniques to accurately understand and model wind dynamics across different locations. The difference in wind speed and the resulting TE behavior between the two stations can be attributed to their geographical locations, demonstrating the significance of coastal and inland dynamics in wind patterns.

4.4.3 Another Example of Transfer Entropy of wind speed

In another example, we compare Stations 50 and 24, as shown in Fig. 4.5. Station 24 is on the coastline at an elevation of 22 meters, while Station 50 is further inland at an elevation of 813.6 meters. The distance between the two stations is approximately 98.5 kilometers.

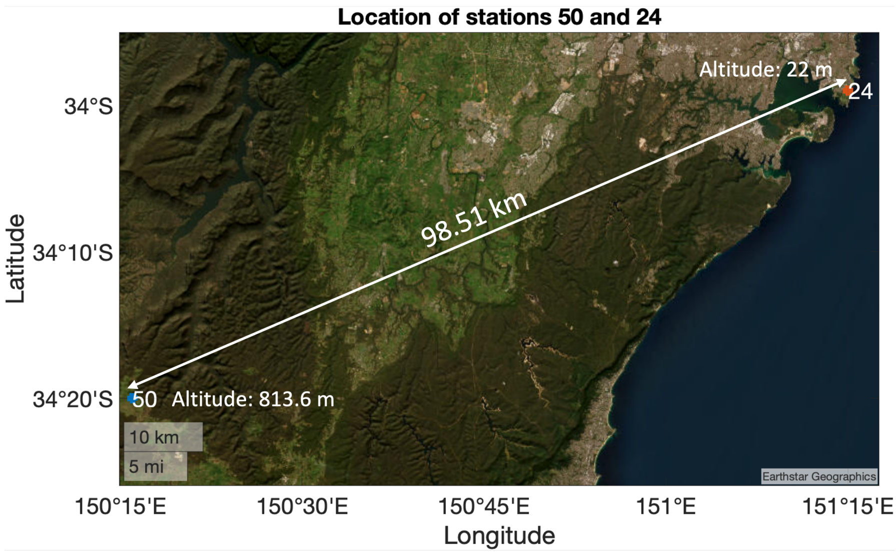


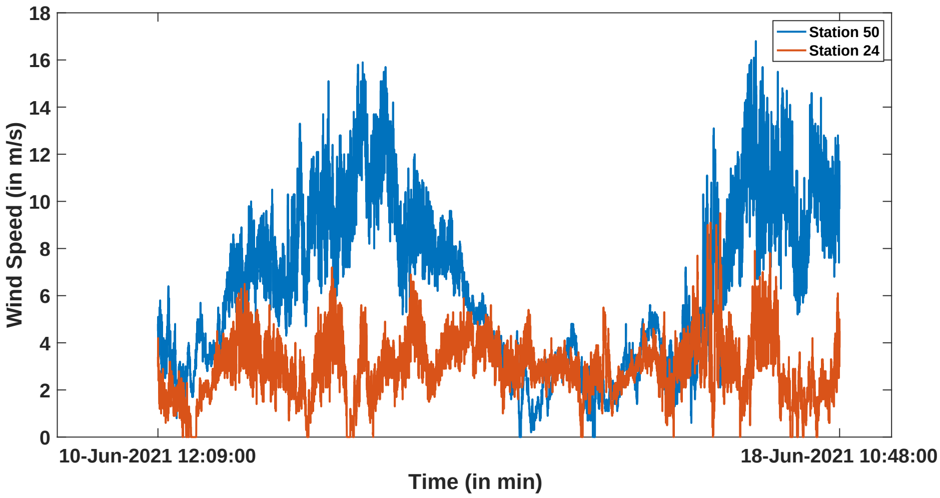
Figure 4.5: Map showing the locations of Stations 50 and 24. Station 24 is on the coastline at an elevation of 22 meters, while Station 50 is further inland at an elevation of 813.6 meters. The distance between the two stations is approximately 98.5 kilometers.

The wind speed data for both stations and the corresponding TE plots are shown in Fig. 4.6. Fig. (4.6a) displays the raw wind speed data for Stations 50 and 24, we observe that the wind speed at Station 50 is higher than at Station 24 during the specific period analyzed. This results in higher TE values from Station 50 to Station 24, as shown in Fig. (4.6b), indicating a stronger influence of the inland station on the coastal station during this period.

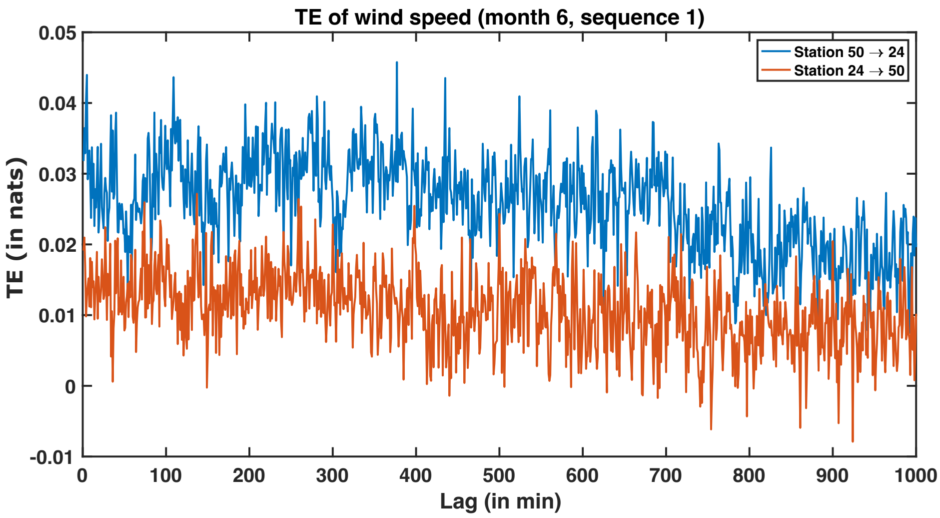
The results highlight an interesting phenomenon: despite Station 24's coastal location, Station 50 exhibits higher wind speeds and a stronger influence on Station 24. This can be attributed to the significant elevation difference between the two stations. Station 50, at 813.6 meters, likely experiences different atmospheric conditions that contribute to stronger winds compared to the lower elevation and coastal influence of Station 24. The higher TE values from Station 50 to Station 24 suggest that during this period, the inland station's wind patterns are exerting a more substantial influence on the coastal station's wind conditions. These observations underscore the complexity of wind dynamics and the importance of considering various factors such as elevation, geographic location, and local atmospheric conditions in wind speed analysis. The contrasting TE behaviors in the two examples demonstrate how different environmental and geographical contexts can influence the direction and strength of wind information transfer between meteorological stations.

4.5 Wind Direction

In this section, instead of analyzing wind speed, we examine the wind direction over the same period between Stations 34 and 24, and compute the Transfer Entropy. Fig. 4.7 presents the wind direction data and the corresponding TE between Stations 34 and 24. The first subplot (Fig. 4.7a) shows the wind direction data, indicating higher variability at Station 34 (inland) compared to Station 24 (coastal). This variability can be attributed to the more complex topographical features and increased surface friction inland. Station 34, located further inland at an elevation of 81.2 meters, experiences wind patterns influenced by surrounding terrain and other geographical factors. In contrast, Station 24, located on the coastline at an elevation of 22 meters, benefits from more stable and consistent wind directions due to the relatively smooth surface of the ocean



(a)



(b)

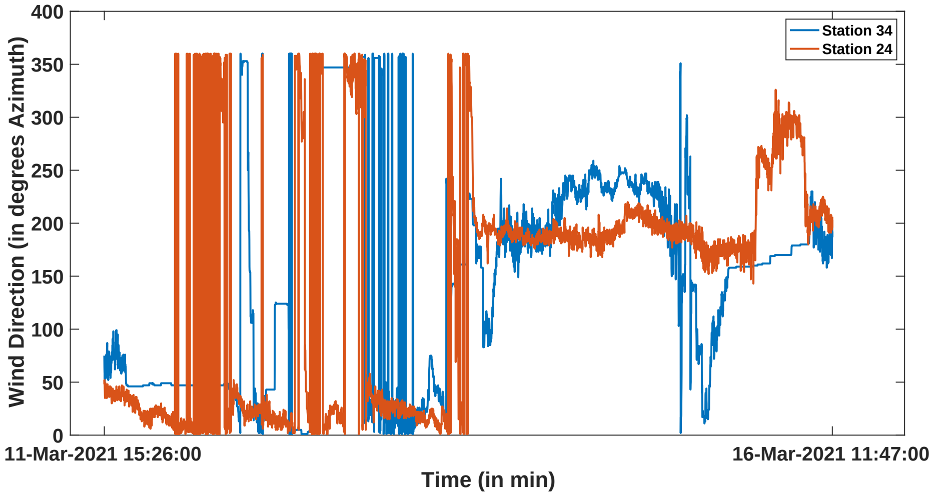
Figure 4.6: (a) Wind speed data for a specific period, showing higher wind speeds at Station 50 (inland) compared to Station 24 (coastal); (b) Transfer Entropy of wind speed for different lags, indicating a stronger influence of Station 50 on Station 24.

and reduced friction. The second subplot (Fig. 4.7b) shows the TE analysis, indicating a stronger influence of Station 34 on Station 24. This suggests that the variable wind directions at the inland station significantly impact the wind patterns at the coastal station. The inland winds, affected by the complex inland geography, seem to carry information that dictates changes in the wind direction at the coastal station. This is particularly noticeable during periods when inland wind variability is high, affecting the coastal wind patterns more substantially.

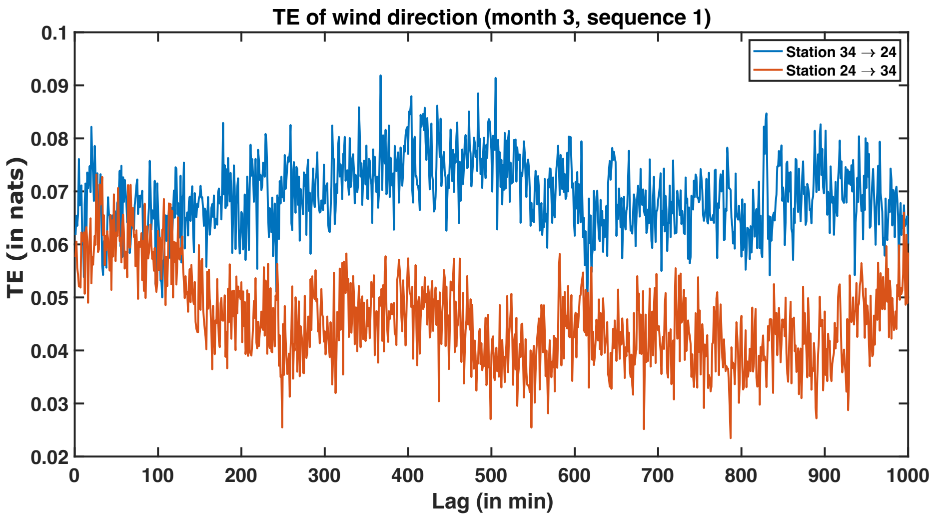
Another example is shown for Stations 50 and 24 in Fig. 4.8. The first subplot (Fig. 4.8a) depicts the wind direction data, demonstrating higher variability at Station 50 (inland) compared to Station 24 (coastal). Station 50, situated at a higher elevation of 813.6 meters and further inland, shows more significant fluctuations in wind direction due to its elevation and surrounding topography. This higher variability in wind direction can cause changes in wind patterns that propagate towards the coastal Station 24. The second subplot (Fig. 4.8b) shows the TE analysis, indicating a stronger influence of Station 50 on Station 24. This is consistent with the previous example, suggesting that inland stations with higher wind direction variability can significantly impact coastal wind patterns. The higher TE values from Station 50 to Station 24 imply that the inland wind dynamics at higher altitudes have a considerable influence on the coastal wind directions.

By comparing the TE of wind speed and wind direction between different pairs of stations, we observe that the influence of one station on another varies depending on the parameter considered. While wind speed TE highlights the stronger influence of coastal stations on inland stations, wind direction TE suggests a significant impact of inland stations on coastal wind patterns. These findings highlight the complexity of wind dynamics and the need for a comprehensive approach to understanding the interactions between different stations.

Further analysis is required to determine which station provides more reliable information about the other. Another approach to consider is analyzing the components of the wind vector and computing the multivariate TE to gain a more detailed understanding of the information transfer between stations.

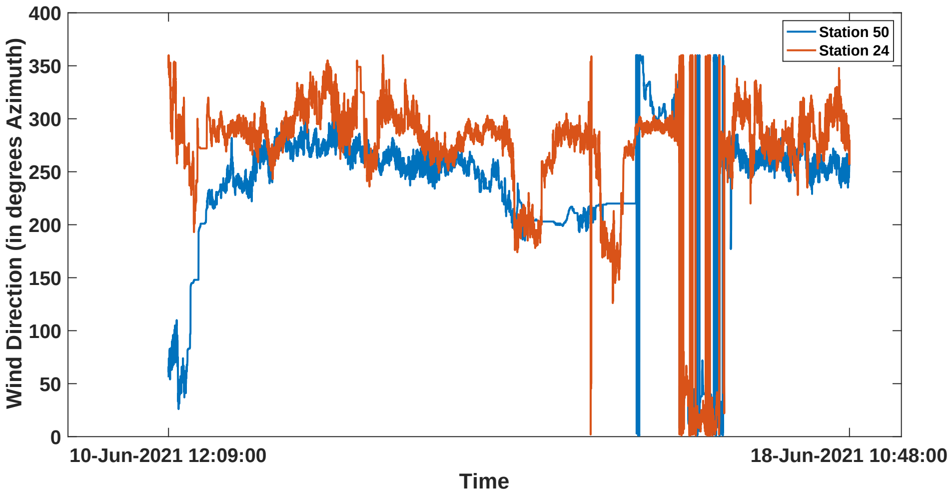


(a)

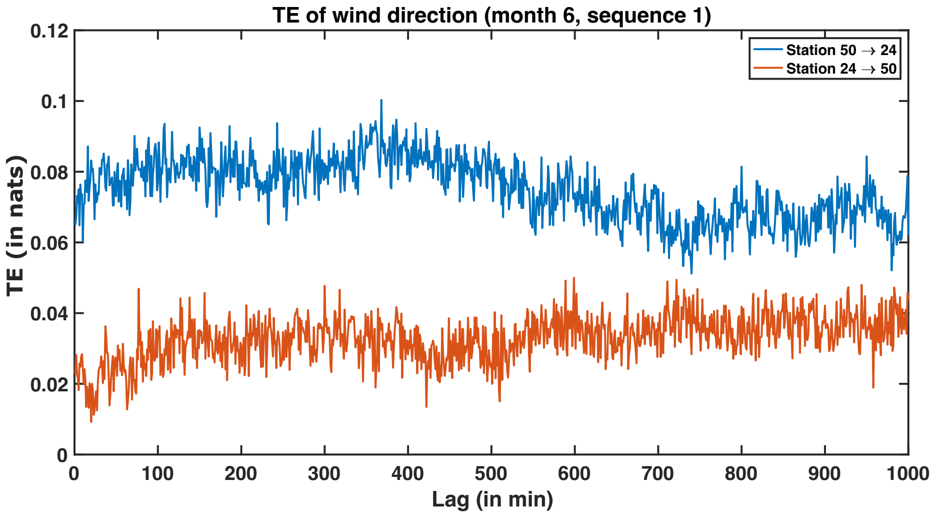


(b)

Figure 4.7: (a) Wind direction data for a specific period, showing higher wind speeds at Station 34 (inland) compared to Station 24 (coastal); (b) Transfer Entropy of wind direction for different lags, indicating a stronger influence of Station 34 on Station 24.



(a)



(b)

Figure 4.8: (a) Wind direction data for a specific period, showing higher wind speeds at Station 50 (inland) compared to Station 24 (coastal); (b) Transfer Entropy of wind direction for different lags, indicating a stronger influence of Station 50 on Station 24.

4.5.1 Wind Rose Analysis

Before delving into multivariate TE, it is essential to analyze the wind roses for the sequences from Stations 34-24 and 50-24. The wind rose is a vital tool in meteorology that shows the distribution of wind speed and direction over a specific period. It provides a comprehensive visualization of how frequently the wind blows from particular directions and at what speeds, which is crucial for understanding wind patterns and their potential influence on different locations.

Fig. 4.9 presents the wind roses for Stations 34 and 24. The first subplot (Fig. 4.9a) depicts the wind rose at Station 34, while the second subplot (Fig. 4.9b) shows the wind rose at Station 24. In these wind roses, Station 34 is positioned in the center as the reference station, while Station 24 is positioned approximately at 101° clockwise from true north with respect to Station 34. The wind roses illustrate the predominant wind directions and their frequencies, offering insights into the wind dynamics between the two stations. Station 34's wind rose shows more variability in wind direction due to its inland location and complex topographical features, whereas Station 24's wind rose reflects more stable and consistent wind patterns typical of coastal regions.

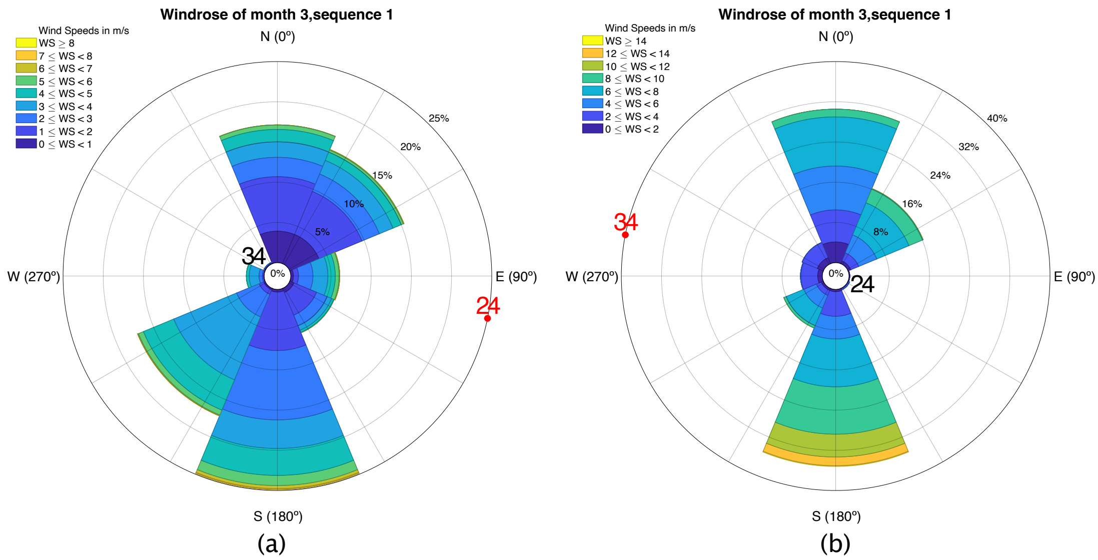


Figure 4.9: Wind roses for Stations 34 and 24 during a specific period: (a) Wind rose at Station 34 (inland); (b) Wind rose at Station 24 (coastal).

Similarly, Fig. 4.10 shows the wind roses for Stations 50 and 24. The first subplot (Fig. 4.10a) illustrates the wind rose at Station 50, and the second subplot (Fig. 4.10b) displays the wind rose at Station 24. Here, Station 50 serves as the reference station in the center, with Station 24 is positioned approximately at

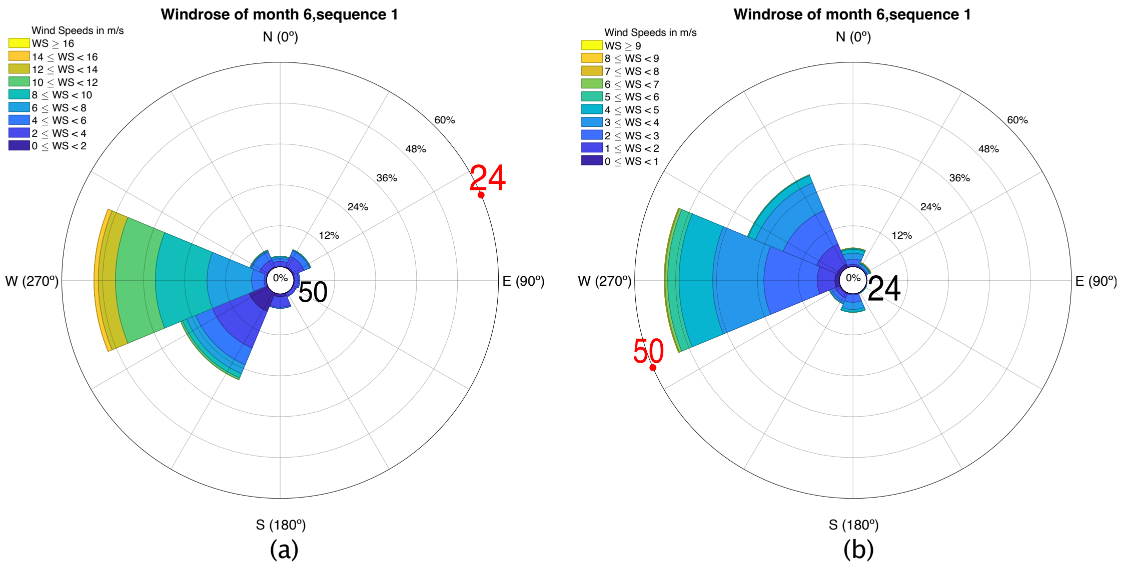


Figure 4.10: Wind roses for Stations 50 and 24 during a specific period: (a) Wind rose at Station 50 (inland); (b) Wind rose at Station 24 (coastal).

67° clockwise from true north with respect to Station 50. These visualizations help in understanding the wind direction distribution and its variability, which is essential for analyzing the TE and wind dynamics between these stations. The wind rose at Station 50 shows significant directional variability due to its higher elevation and inland location, while Station 24's wind rose remains consistent with its coastal characteristics.

The wind roses provide valuable context for understanding the interactions between stations. They help visualize the directional characteristics of wind patterns, showing how often winds blow from specific directions and at what speeds. This information is critical for analyzing the TE between stations, as it highlights the dominant wind directions and their potential influence on nearby stations.

4.6 Multivariate Transfer Entropy Analysis

To gain a deeper understanding of wind dynamics and the information flow between stations, we now consider the multivariate Transfer Entropy using the x and y components of the wind vector (Eq. 4.2), as described in Section 4.3.2.

This approach captures the combined influence of wind speed and direction, providing a more detailed analysis of directional information flow between stations.

Fig. 4.11 illustrates a part of the sequence for the x and y wind components at Station 34. This figure demonstrates the variability in wind speed and direction, highlighting the complexity of wind patterns. By analyzing these components, we gain insights into the bidirectional influence between stations that are not captured when considering wind speed or direction alone. In this figure, the

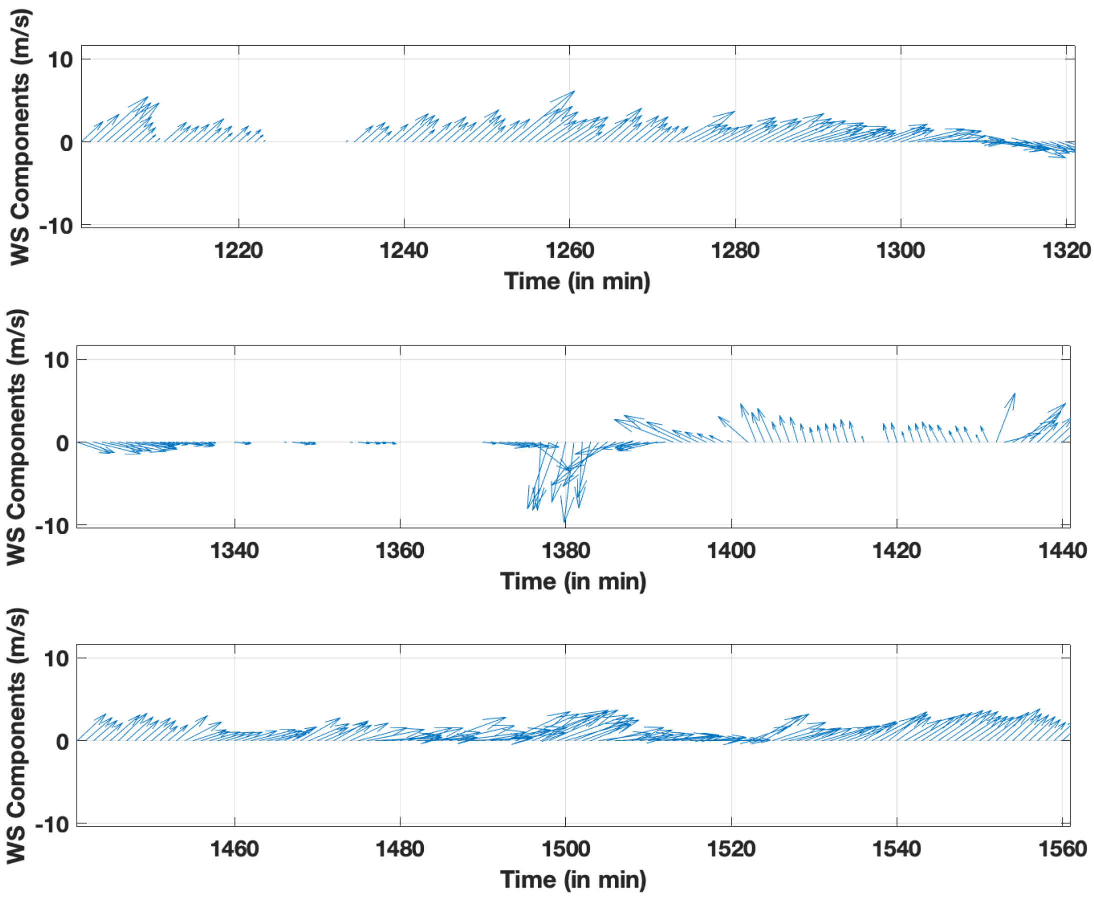


Figure 4.11: Wind components at Station 34 for a specific period, showing the variability in wind speed and direction.

x and y components represent the horizontal and vertical components of the wind vector, respectively. A positive x component indicates wind blowing from the west, while a negative x component indicates wind blowing from the east.

Similarly, a positive y component indicates wind blowing from the south, and a negative y component indicates wind blowing from the north. The arrows in the figure indicate both the direction and magnitude of the wind vector at different times, emphasizing the dynamic nature of wind flow at the station.

Previous analyses of TE for wind speed and wind direction alone have provided valuable insights but are insufficient to capture the full complexity of wind dynamics. These univariate analyses do not fully encompass the intricate interactions between wind speed and direction components. To address this limitation, we now conduct a multivariate TE analysis using the x and y components of the wind vector. This multivariate approach considers the joint influence of both components, providing a more comprehensive understanding of the information transfer between stations. By analyzing the multivariate TE, we can identify how the combined effects of wind speed and direction contribute to the overall wind dynamics, offering a richer and more detailed perspective on the interactions between stations.

Figs. 4.12 and 4.13 present the multivariate TE results for Stations 34-24 and 50-24, respectively. In Fig. 4.12, the TE from Station 34 to Station 24 is higher than the TE from Station 24 to Station 34, indicating a dominant influence of the inland station over the coastal station. The TE from Station 34 to Station 24 shows a maximum value at a lag of 438 minutes. This peak indicates that the information transfer from Station 34 to Station 24 is most significant at this lag. By lagging Station 34's wind data by 438 minutes, we observe that the most substantial amount of information is transferred to Station 24, highlighting the influence of the inland station on the coastal station's wind dynamics. Additionally, searching for lags that show a peak value in the TE can help identify the points where information transfer is maximized between the two stations, suggesting that significant events or changes in wind patterns might have occurred at these specific lags. The results for Station 50 to Station 24 in Fig. 4.13 also highlight important information transfer patterns, further emphasizing the role of inland stations in influencing coastal wind dynamics.

The multivariate TE analysis provides a comprehensive understanding of wind dynamics by considering the joint effects of wind speed and direction. This approach uncovers intricate patterns of information transfer between stations

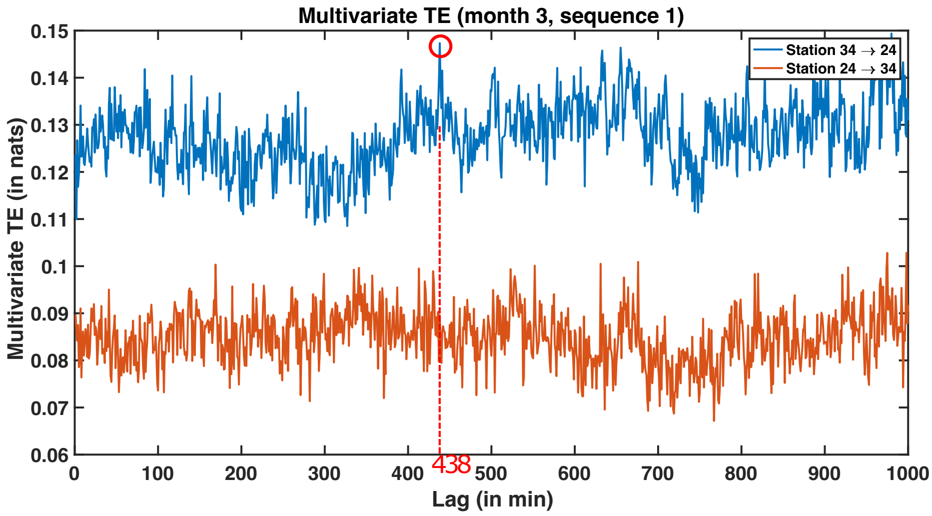


Figure 4.12: Multivariate TE between Stations 34 and 24 for a specific period, highlighting a maximum TE at a lag of 438 minutes.

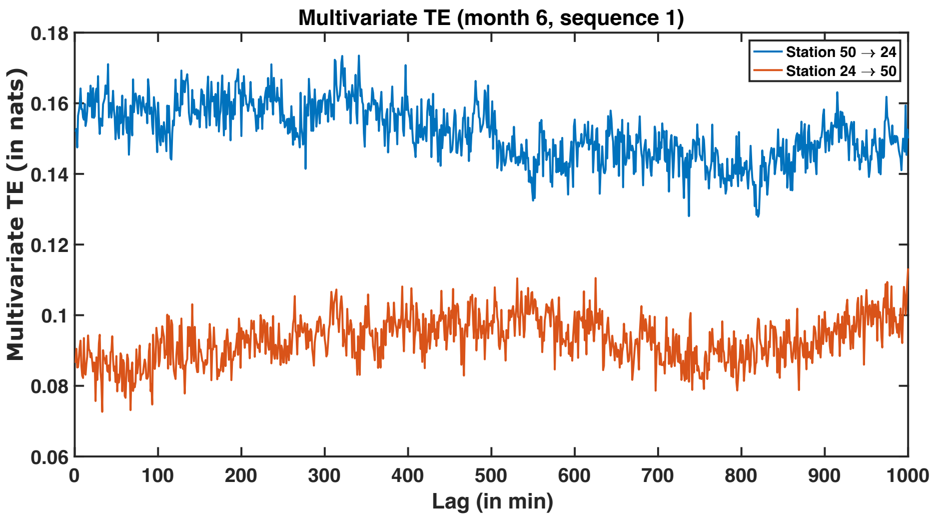


Figure 4.13: Multivariate TE between Stations 50 and 24 for a specific period.

that are not evident from univariate analyses alone, thus offering a more nuanced perspective on the interactions between different wind components across various locations.

4.7 Conclusion

In this chapter, we applied and validated the Multivariate Transfer Entropy methodology to analyze wind speed and direction data from multiple meteorological stations in Sydney, Australia. Our primary objective was to uncover the directional flow of information within the wind data, enhancing our understanding of wind dynamics and improving predictive models for weather forecasting and renewable energy management. By transforming the wind data into x and y components, we enabled a detailed multivariate analysis that captured the complex interactions between wind speed and direction. The use of both univariate and multivariate TE provided a comprehensive view of how information propagates between different meteorological stations.

Our analysis revealed several key insights, including the observation that coastal stations exert a stronger influence on inland stations, likely due to the consistent and robust sea breeze effects that dominate coastal wind patterns. Additionally, the multivariate TE approach uncovered intricate patterns of information transfer that were not apparent in the univariate analysis, demonstrating the value of considering wind speed and direction jointly. The methodology for calculating Multivariate TE, rigorously tested in Appendix B using coupled Rössler systems, proved to be robust and reliable for analyzing the complex dynamics of wind data. This validation underscores the effectiveness of our approach in capturing the directional dependencies and nonlinear interactions inherent in atmospheric processes.

In summary, the integration of univariate and multivariate TE analyses has provided a deeper understanding of wind dynamics, highlighting the intricate information flow between meteorological stations. This chapter has demonstrated the applicability and strength of our TE methodology in a real-world context, setting the stage for using TE to forecast wind, which is part of our future perspectives.

Chapter 5

Horizon Estimation of Chaotic Systems

Outline of the current chapter

In this chapter, we discuss the Nonlinear Local Lyapunov Exponent (NLLE) method, which is used to estimate the predictability time of chaotic systems. We focus on the application of this method to the Lorenz63 system, a well-known chaotic system. This chapter introduces an alternative approach to the Nonlinear Local Lyapunov Exponent (NLLE) algorithm, enhancing its applicability for studying the predictability of complex spatiotemporal systems. Future work will expand on this method to address the predictability in more complex systems.

5.1 Introduction

The predictability time of a chaotic system is a crucial measure that indicates how long into the future we can accurately predict the state of the system. This is especially important in fields such as meteorology, finance, and any other domain where chaotic dynamics play a significant role. The maximum predictability time, also known as the predictability horizon, represents the upper limit beyond which the system's future state becomes effectively unpredictable due to the exponential growth of small perturbations. Chaotic systems, by their nature, exhibit sensitive dependence on initial conditions, meaning that tiny differences in starting states can lead to vastly different outcomes. This inherent unpredictability is quantified by the Lyapunov exponent, which measures the rate at which nearby trajectories diverge. The theoretical Lyapunov time, which is inversely proportional to the largest Lyapunov exponent, provides an estimate of the timescale over which the system's predictability degrades.

The calculation of Lyapunov exponents from a time series has been a fundamental tool in the study of chaotic systems. Wolf et al. (1985) [35] introduced a method that involves reconstructing phase space using time delay embedding, which is crucial for analyzing chaotic time series data. However, this method is sensitive to the selection of embedding parameters and requires a large amount of data for accurate computation, especially for high-dimensional systems like the atmosphere. Chen et al. (2006) described a preliminary algorithm to estimate the NLLE and applied it to study the predictability of the 500-hPa geopotential height field [198]. Their work provided a foundation for subsequent methods that aimed to improve the accuracy and applicability of NLLE calculations in various chaotic systems. Building on these foundational works, Li and Ding (2011) developed an advanced algorithm to quantify the predictability limits of chaotic systems more accurately [199]. Their method focuses on local dynamics, providing a more nuanced and precise picture of predictability. Unlike traditional methods that might rely on global averages or linear approximations, the NLLE method by Li and Ding offers better insights into the local stability of trajectories in chaotic systems. The key idea behind their method is to measure how small differences in initial conditions evolve over time, thereby estimating the rate at which errors grow. This rate is quantified by the Lyapunov exponents, with the NLLE specifically focusing on local and nonlinear aspects of this growth. This method not only provides a quantitative estimate of the predictability limit but also offers insights into the local dynamics that drive error growth.

5.2 The NLLE Method

The NLLE method provides a nonlinear approach to quantifying the predictability limit of chaotic systems by focusing on the local error growth rate without linearizing the governing equations. Traditional methods, such as those based on the global Lyapunov exponent, often fall short in accurately capturing the complex, local dynamics of chaotic systems. The NLLE measures the average growth rate of initial errors in nonlinear dynamical models. It has been effectively applied to a wide range of chaotic systems, from simple models like the Lorenz63 to more complex, high-dimensional systems such as the atmosphere. By identifying local dynamical analogs from historical data, the NLLE method allows for a detailed analysis of error growth and predictability in specific regions and under specific conditions. This method not only provides a quantitative estimate of the predictability limit but also offers insights into the local dynamics that drive error growth, making it a valuable tool for both theoretical studies and practical applications in chaotic systems.

5.3 Estimating the Maximum Predictability Time of the Lorenz63 System

In this section, we aim to estimate the maximum predictability time of a chaotic system, specifically the Lorenz63 system, using the NLLE algorithm. The Lorenz63 system, uncovered by Edward Lorenz in 1963 during his exploration of atmospheric convection, stands as a seminal example within chaos theory [1].

The governing equations of the Lorenz63 system are:

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \beta z, \end{cases} \tag{1.3}$$

where $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ represent the system's parameters, linked to the Prandtl number, Rayleigh number, and specific physical dimensions of the convection box, respectively. We have discussed the Lorenz system in detail in Chapter 1 (Section 1.1.4).

The theoretical Lyapunov time, which is inversely proportional to the largest Lyapunov exponent, provides an estimate of the timescale over which the system's predictability degrades. In the context of the Lorenz system, the maximum Lyapunov exponent λ_{max} is approximately 0.91, leading to a theoretical Lyapunov time $\tau_p = \lambda_{max}^{-1} = 1.1$.

5.3.1 Steps in Estimating the Maximum Predictability Time using NLLE

The objective of this analysis is to estimate the maximum predictability time of the Lorenz63 system using the NLLE algorithm. The procedure involves the following steps:

Step 1: Seeking Local Dynamical Analogs

To begin, select a reference point $x(t_0)$. For each reference point, consider all potential analogous points $x(t_j)$ within a short time interval. According to Li and Ding (2011) [199], this short time interval is the time taken for the autocorrelations of the variable to drop to 0.9, ensuring that a good analog pair is not merely due to persistence. For our analysis, we use a similar criterion. In our method, we identify the analogous points using the 'findsignal' function in MATLAB, which is designed to find segments within the data that match a given reference signal. This approach allows for a more automated and accurate identification of local dynamical analogs. First, calculate the initial distance between the two points: $d_i = |x(t_0) - x(t_j)|$. Next, evaluate the evolutionary distance between their trajectories over a short initial period to account for their dynamic similarity:

$$d_e = \sqrt{\frac{1}{K+1}\sum_{i=0}^{K}[x(t_i) - x(t_{j+i})]^2}$$

The total distance is then given by: $d_t = d_i + d_e$. If d_t is sufficiently small, the two points are considered locally dynamically analogous, indicating that they will exhibit similar behavior over the short term. In our method, the difference lies in how we identify these analogous points by adjusting the length of the signal we used in the 'findsignal' function to better capture the local dynamics.

Step 2: Trajectory Analysis

For each pair of analogous points identified in Step 1, analyze their trajectories. According to Li and Ding (2011) [199], this comparison is made over a time period defined by a certain number of steps K such that $K\Delta t$ represents a short initial period. At time $t_i = t_0 + i\Delta$, compare the positions along the reference trajectory $x(t_i)$ and the analogous trajectory $x(t_{k+i})$. Calculate the initial distance between the two points at the start of the analysis: $L(t_0) = |x(t_0) - x(t_k)|$. Then, determine the distance between the trajectories at time t_i: $L(t_i) = |x(t_i) - x(t_{k+i})|$. The growth rate of the initial error during the time interval $(t_i - t_0)$ is calculated as:

$$\xi_1(t_i) = \frac{1}{t_i - t_0} \ln \frac{L(t_i)}{L(t_0)} \tag{5.1}$$

This quantifies how quickly small differences in initial conditions grow over time, reflecting the system's sensitivity to initial perturbations.

Example of Analog Trajectory Identification: To illustrate the process of identifying analogous trajectories, consider Fig. 5.1, which presents the Lorenz X variable along with the reference and analog trajectories. The top part of the figure displays the overall behavior of the Lorenz X variable, the reference trajectory, and the analog trajectory over time, providing a comprehensive overview. The bottom part zooms in on the reference and analog trajectories for a clearer comparison, demonstrating how dynamically similar points are identified. This example showcases the method's effectiveness in finding and comparing trajectories within the Lorenz system.

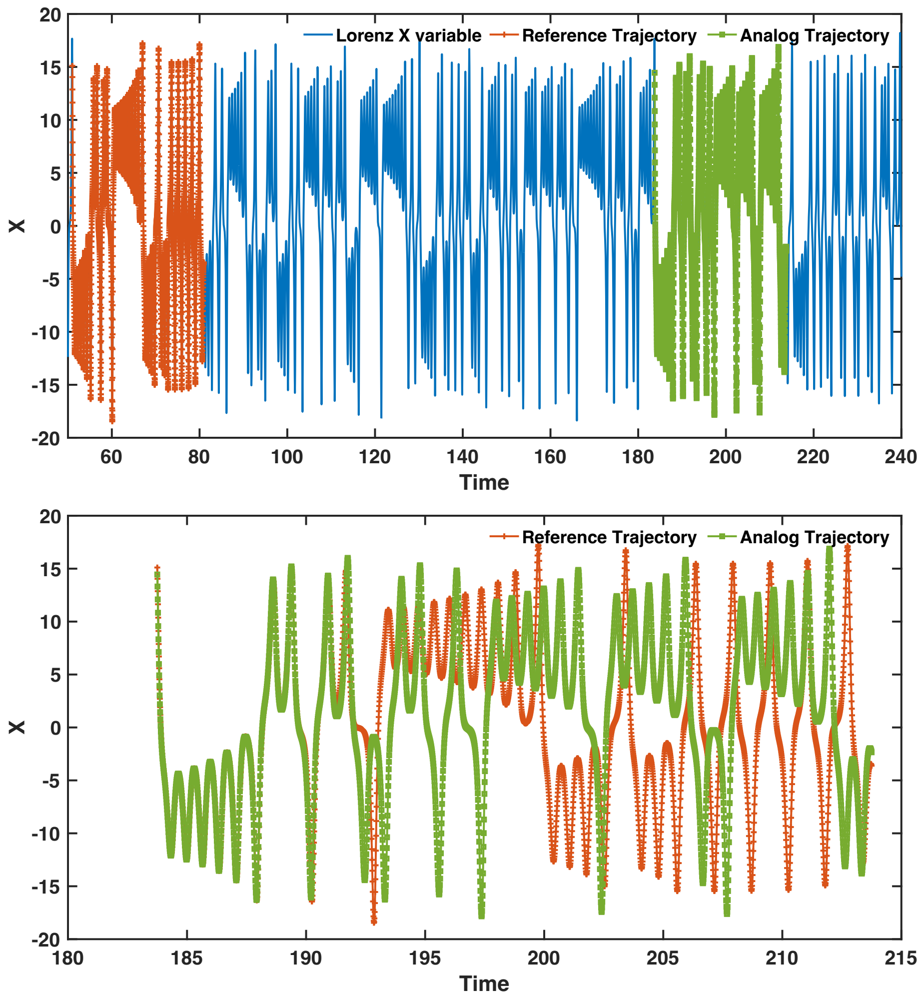


Figure 5.1: Identification of analogous trajectories in the Lorenz X variable. The top plot shows the Lorenz X variable (blue), the reference trajectory (red), and the analog trajectory (green). The bottom plot provides a zoomed-in view of the reference and analog trajectories for better comparison.

Step 3: Choosing Another Reference State

Select another reference state $x(t_1)$ and repeat steps 1 and 2. This process should be repeated for a sufficiently large number of reference points to ensure a comprehensive analysis of the system's behavior across different regions of the phase space.

Step 4: Averaging Error Growth Rates

Repeat the analysis for all selected reference points N. The mean NLLE is obtained by averaging the error growth rates calculated for each reference point:

$$\bar{\xi}(t_i) = \frac{1}{N}\sum_{k=1}^{N} \xi_k(t_i) \tag{5.2}$$

This provides a robust measure of the average local Lyapunov exponent, reflecting the typical rate of error growth in the system.

Step 5: Estimating the Relative Growth of Initial Errors

From Eqs. 5.1 and 5.2, we obtain the approximation of the Relative Growth of the Initial Error (RGIE):

$$\bar{\Phi}(t_i) = \exp[\bar{\xi}(t_i)(t_i - t_0)] \tag{5.3}$$

This function describes how the initial error evolves over time, providing insight into the system's predictability.

5.3.2 Impact of Parameter Variations on Predictability Time Estimation

Through our analysis, we observed that varying the parameter representing the correlation length did not significantly change the estimation of the maximum predictability time. This indicates that the selection of local dynamical analogs is relatively robust to the exact choice of correlation length. However, changing the parameter representing the initial length of the trajectory segment had a substantial impact on the estimation. Increasing the initial segment length resulted in a higher predictability horizon and lower RGIE. This suggests that using a longer initial segment allows for more accurate identification of dynamically similar points, thereby reducing the initial error growth rate and increasing the estimated predictability time.

For instance, as shown in Fig. 5.2, when using a smaller initial segment length, the saturation point is clearer, making it easier to identify the prediction horizon.

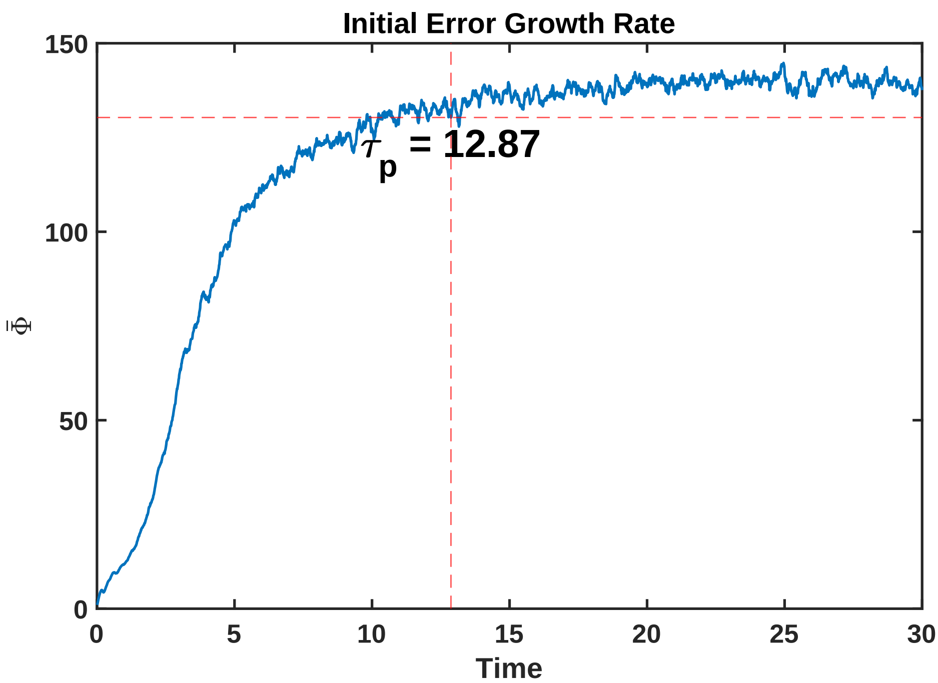


Figure 5.2: Initial Error Growth Rate with a smaller initial segment length. The figure shows the evolution of the RGIE with a clear saturation point but higher initial error growth.

However, this comes with a higher initial error growth rate, which might reduce the accuracy of the analog selection over time. In contrast, as shown in Fig. 5.3, a larger initial segment length results in lower initial error growth, indicating more accurate analog identification, although the saturation point becomes less distinct.

Given these observations, using a longer initial segment length is preferable for a more accurate estimation of the maximum predictability time, as it provides a more precise selection of dynamically similar points, despite the less distinct saturation point.

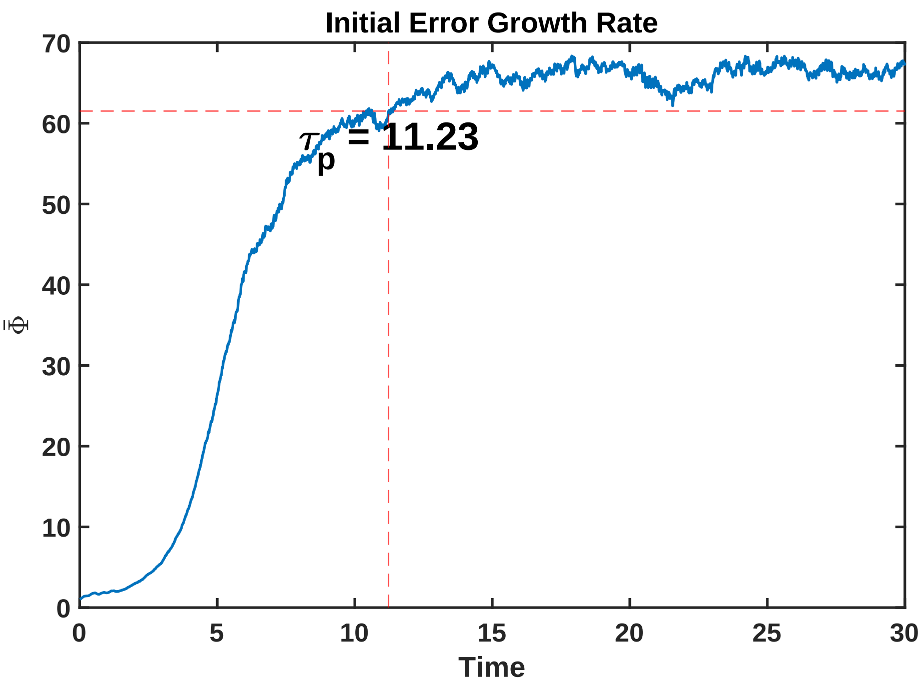


Figure 5.3: Initial Error Growth Rate with a larger initial segment length. The figure shows the evolution of the RGIE with a less distinct saturation point but lower initial error growth.

These results highlight the importance of choosing an appropriate initial segment length for accurate predictability analysis using the NLLE method. The prediction horizon time τ_p can be defined as the time at which the evolution of $\bar{\Phi}(t_i)$ with increasing t_i reaches 90% of its saturation value. For the variable X, the nonlinear prediction horizon $\tau_{p_{\text{nonlinear}}}$ varies between 11 and 13 for different initial segment lengths. Additionally, this method demonstrates that the prediction horizon is larger than that indicated by the theoretical Lyapunov time $\tau_p = 1.1$.

5.4 Conclusion

In this chapter, we presented a modified NLLE method to estimate the maximum predictability time of chaotic systems, specifically the Lorenz63 system. We introduced an alternative approach to the NLLE algorithm that involves using the findsignal function in MATLAB for identifying local dynamical analogs. Our analysis showed that varying the initial length of the trajectory segment significantly impacts the estimation of the predictability horizon. We found that a longer initial segment length results in a higher predictability horizon and lower initial error growth, despite a less distinct saturation point. This method provides a more accurate estimation of the maximum predictability time, demonstrating a predictability horizon larger than that indicated by the theoretical Lyapunov time. Future work will extend this method to study the predictability of complex spatiotemporal systems.

Conclusion

The primary objective of this thesis was to improve the prediction of extreme events in chaotic systems, with a particular focus on optical systems modeled by the Lugiato-Lefever equation. We aimed to extend the predictability horizon of these systems using a combination of machine learning techniques and information theory methods, particularly transfer entropy. The methodologies developed were also tested on real-world wind dynamics, showcasing their broader applicability beyond controlled optical systems.

At the outset, we explored the foundations of chaotic systems, especially spatiotemporal chaos, where complex patterns unfold over both time and space. Understanding these dynamics was essential for studying extreme events in optical systems. In chaotic systems, small initial changes can rapidly lead to significantly different outcomes, making long-term prediction a challenging endeavor. Nevertheless, recent advances in machine learning, particularly neural networks, have provided new tools for identifying patterns in these complex systems. Our investigation aimed to leverage these tools to make predictions in such highly unpredictable environments. In the context of optical fiber ring resonators, we applied machine learning models to forecast the full chaotic dynamics. While these models performed well for short-term predictions, they struggled with long-term forecasts due to the extensive chaotic nature of the system. As the complexity of the system increased, the models' ability to capture the dynamics declined. This led us to explore the integration of transfer entropy—a method from information theory that measures the directional flow of information between different parts of the system. By incorporating transfer entropy, we shifted from attempting to predict the entire system's behavior to focusing on localized regions where extreme events were likely to occur. This hybrid approach allowed for better accuracy in identifying early signs of these events, thereby improving the overall forecasting capability.

A significant discovery in this thesis was the detailed analysis of the transition from soliton-based dynamics to spatiotemporal chaos and fully developed turbulence in a bistable optical system governed by the Lugiato-Lefever equation (LLE). By progressively increasing the pump parameter, we observed a shift from a regime dominated by solitons to one characterized by phase turbulence. Our analysis demonstrated self-similar behavior in both spatial and temporal domains, with power-law spectra indicative of turbulent-like dynamics, particularly in the phase. However, the amplitude spectra decayed more steeply, suggesting different underlying mechanisms. This was explained by the identification of stable linear waves propagating along the homogeneous steady-state (HSS) solution. These waves play a crucial role in initiating the energy cascade in the phase spectra, as confirmed through the dispersion relation analysis. Further analysis using two-point correlation functions showed that, outside the correlation domain of localized structures, the fluctuations spectrum follows a power-law decay. Furthermore, the application of structure functions enabled us to characterize higher-order correlations, revealing the onset of intermittency and the divergence from the system's scaling behavior at larger scales. The emergence of topological defects marked another significant finding of this chapter and serves as an indicator of turbulence. Overall, the analysis in this chapter has enriched our comprehension of the transition to turbulence in such systems.

Extending the insights gained from optical systems, we extended our methods to the analysis of real-world wind dynamics by applying univariate and multivariate Transfer Entropy (TE) to wind speed and direction data from meteorological stations. This analysis revealed the directional flow of information between stations, uncovering critical directional dependencies and complex interactions. By transforming wind data into x and y components, we captured dynamics that would have been overlooked using univariate approaches alone. Our findings highlighted the significant influence of coastal stations on inland areas, largely due to the dominant sea breeze effects. The robustness of our TE methodology was further validated using coupled Rössler systems.

Finally, we addressed one of the key challenges in forecasting chaotic systems:

the predictability horizon. By employing the Nonlinear Local Lyapunov Exponent (NLLE) method, we estimated how far into the future reliable predictions could be made before the chaotic nature of the system rendered them inaccurate. This analysis not only shed light on the fundamental limitations of predicting chaotic systems but also pointed toward potential methods for extending the predictability horizon.

In conclusion, while long-term prediction of chaotic systems remains inherently challenging due to their unpredictable nature, this thesis demonstrates that by integrating machine learning with information theory, it is possible to significantly improve forecast accuracy and extend the prediction horizon. This hybrid approach of focusing on local regions through transfer entropy offers a practical pathway to anticipate extreme events. Our findings contribute to both the fundamental understanding of chaotic systems and the development of practical tools for forecasting extreme events, whether in optical systems or natural phenomena like wind dynamics.

While this thesis has made significant progress in forecasting extreme events in chaotic systems, there is still much room for improvement. Future work should focus on refining machine learning models to better capture the long-term behavior of chaotic systems. This could involve developing more advanced architectures, optimizing data handling techniques, and improving the training process to better capture the intricacies of chaotic behavior. Furthermore, exploring new algorithms and hybrid models that combine the strengths of various machine learning techniques could help overcome the current limitations of long-term predictions.

Another promising direction for future research involves the integration Transfer Entropy with machine learning models to enhance wind dynamics forecasting. By combining these methods, we could identify key stations that hold the most predictive power over others, optimizing the selection of stations for forecasting. This approach would refine predictions by focusing on the most critical information pathways, further extending the predictability horizon. Applying this combined methodology across diverse meteorological contexts could significantly improve resource management, particularly in fields like renewable energy, where accurate wind forecasts are essential.

Appendix A

Linear Stability Analysis

Outline of the current chapter

A.1 Introduction

Studying the stability analysis of the Homogeneous Steady State (HSS) solution of the Lugiato-Lefever Equation (LLE) can be performed by introducing the complex conjugate of Eq. 1.12 (Here Eq. A.1). This analysis helps in understanding how small perturbations around the steady state evolve over time, which is crucial for determining the stability of the system.

$$\frac{\partial \psi}{\partial t} = S - (1 + i\Delta)\psi + i\frac{\partial^2 \psi}{\partial \tau^2} + i \mid \psi \mid^2 \psi \tag{A.1}$$

$$\frac{\partial \psi^*}{\partial t} = S - (1 + i\Delta)\psi^* + i\frac{\partial^2 \psi^*}{\partial \tau^2} + i \mid \psi^* \mid^2 \psi^* \tag{A.2}$$

A.2 Perturbation Analysis

To analyze the stability, we introduce perturbations around the steady state solution ψ_0. Let $\psi = \psi_0 + \delta\psi$ and $\psi^* = \psi_0^* + \delta\psi^*$. Here, $(\delta\psi, \delta\psi^*)$ represent small perturbations around the stationary state (ψ_0, ψ_0^*). Substituting these into Eqs. A.1 and A.2 yields:

$$\frac{\partial \psi_0}{\partial t} + \frac{\partial \delta\psi}{\partial t} = S - (1 + i\Delta)\psi_0 - (1 + i\Delta)\delta\psi + i\frac{\partial^2 \psi_0}{\partial \tau^2} + i\frac{\partial^2 \delta\psi}{\partial \tau^2} + i \mid \psi_0 \mid^2 (\psi_0 + 2\delta\psi + \delta\psi^*) \tag{A.3}$$

$$\frac{\partial \psi_0^*}{\partial t} + \frac{\partial \delta\psi^*}{\partial t} = S - (1 + i\Delta)\psi_0^* - (1 + i\Delta)\delta\psi_0 - i\frac{\partial^2 \psi_0^*}{\partial \tau^2} - i\frac{\partial^2 \delta\psi^*}{\partial \tau^2} - i \mid \psi_0 \mid^2 (\psi_0 + \delta\psi + 2\delta\psi^*) \tag{A.4}$$

By eliminating the homogeneous solution, we get:

$$\frac{\partial \delta\psi}{\partial t} = -(1 + i\Delta)\delta\psi + i\frac{\partial^2 \delta\psi}{\partial \tau^2} + 2i \mid \psi_0 \mid^2 \delta\psi + i \mid \psi_0 \mid^2 \delta\psi^* \tag{A.5}$$

$$\frac{\partial \delta\psi^*}{\partial t} = -(1 + i\Delta)\delta\psi_0 - i\frac{\partial^2 \delta\psi^*}{\partial \tau^2} - 2i \mid \psi_0 \mid^2 \delta\psi^* - i \mid \psi_0 \mid^2 \delta\psi^* \tag{A.6}$$

Which can be written in matrix form as:

$$\frac{\partial}{\partial t}\begin{pmatrix} \delta\psi \\ \delta\psi^* \end{pmatrix} = \begin{pmatrix} -1 - i(\Delta - 2 \mid \psi_0 \mid^2 - \frac{\partial^2}{\partial \tau^2}) & i \mid \psi_0 \mid^2 \\ -i \mid \psi_0 \mid^2 & -1 + i(\Delta - 2 \mid \psi_0 \mid^2 - \frac{\partial^2}{\partial \tau^2}) \end{pmatrix}\begin{pmatrix} \delta\psi \\ \delta\psi^* \end{pmatrix} \tag{A.7}$$

We add a small perturbation of the form:

$$\begin{pmatrix} \delta\psi \\ \delta\psi^* \end{pmatrix} = \begin{pmatrix} \delta\psi_0 \\ \delta\psi_0^* \end{pmatrix} e^{\lambda t + i\omega\tau} \tag{A.8}$$

where $(\delta\psi_0, \delta\psi_0^*)$ represents the initial amplitude of the perturbation and is small compared to the stationary state (ψ_0, ψ_0^*). The linearized problem becomes:

$$\lambda\begin{pmatrix} \delta\psi_0 \\ \delta\psi_0^* \end{pmatrix} = \begin{pmatrix} -1 - i(\Delta - 2 \mid \psi_0 \mid^2 + \omega^2) & i \mid \psi_0 \mid^2 \\ -i \mid \psi_0 \mid^2 & -1 + i(\Delta - 2 \mid \psi_0 \mid^2 + \omega^2) \end{pmatrix}\begin{pmatrix} \delta\psi_0 \\ \delta\psi_0^* \end{pmatrix} \tag{A.9}$$

There are solutions $(\delta\psi_0, \delta\psi_0^*) \neq (0, 0)$ if and only if:

$$\det\begin{pmatrix} -(1 + \lambda) - i(\Delta - 2 \mid \psi_0 \mid^2 + \omega^2) & i \mid \psi_0 \mid^2 \\ -i \mid \psi_0 \mid^2 & -(1 + \lambda) + i(\Delta - 2 \mid \psi_0 \mid^2 + \omega^2) \end{pmatrix} = 0 \tag{A.10}$$

This yields the following dispersion relation:

$$(\lambda + 1)^2 + (\Delta - 2I_s + \omega^2)^2 - I_S^2 = 0 \tag{A.11}$$

The eigenvalues are obtained by solving this relation:

$$\lambda_{\pm} = -1 \pm \sqrt{I_S^2 - (\Delta - 2I_s + \omega^2)^2} \tag{A.12}$$

A.3 Eigenvalue Analysis

The stationary solution is unstable if one of the roots has a positive real part, which is the growth rate in amplitude of the perturbation. Hence, λ_- is always negative, and only λ_+ contributes to the instability of the stationary solution.

First Derivative of λ_+:

$$\frac{\partial \lambda_+}{\partial \omega} = \frac{-2\omega(\omega^2 + \Delta - 2I_s)}{\sqrt{I_S^2 - (\omega^2 + \Delta - 2I_s)^2}} \tag{A.13}$$

Setting $\frac{\partial \lambda_+}{\partial \omega} = 0$ gives the following solutions:

$$\begin{cases} \omega_c^{(0)} &= 0 \\ \omega_c^2 &= 2I_s - \Delta \end{cases}$$

Second Derivative of λ_+:

$$\frac{\partial^2 \lambda_+}{\partial \omega^2} = \frac{-2[I_s^2(3\omega^2 + \Delta - 2I_s) - (\omega^2 + \Delta - 2I_s)^3]}{\sqrt{I_s^2 - (\omega^2 + \Delta - 2I_s)^2}^{\,3}} \tag{A.14}$$

a) If $\omega = \omega_c^{(0)}$:

$$\frac{\partial^2 \lambda_+}{\partial \omega^2}(0) = \frac{-2(\Delta - 2I_s)}{\sqrt{I_S^2 - (\Delta - 2I_s)^2}}$$

$\omega_c^{(0)} = 0$ is a maximum if $\frac{\partial^2 \lambda_+}{\partial \omega^2}(0) < 0$, that is, if $\Delta > 2I_s$ and $I_S^2 - (\Delta - 2I_s)^2 > 0$. This is satisfied for $\frac{\Delta}{3} \leq I_s \leq \Delta$. Finally, $\frac{\Delta}{3} \leq I_s \leq \frac{\Delta}{2}$.

b) If $\omega = \omega_c^2$:

$$\frac{\partial^2 \lambda_+}{\partial \omega^2}(\omega_c^2) = -4I_s(2I_s - \Delta)$$

$\omega_c^2 = 2I_s - \Delta$ is a maximum if $\frac{\partial^2 \lambda_+}{\partial \omega^2}(\omega_c^2) < 0$, that is, if $\Delta < 2I_s$.

The threshold $\lambda_+(\omega_c) = 0$:

a) Case $\omega_c^{(0)} = 0$:

$$I_\pm = \frac{2\Delta + \sqrt{\Delta^2 - 3}}{3}$$

if $\Delta \geq \sqrt{3}$. This is the same expression for the two turning points of the stationary solution curve.

b) Case $\omega_c^2 = 2I_s - \Delta$:

$$I_m = 1$$

Based on the expression of λ_+ (Eq. A.12), it is important to study the instability of the system by distinguishing between the two characteristics of instability: the homogeneous one corresponding to $\omega_c^{(0)}$, denoted $\omega_c^{(1)}$, and the Turing instability at $\omega_c = \sqrt{2I_s - \Delta}$, denoted $\omega_c^{(2)}$.

In the monostable case, $\Delta < \sqrt{3}$: Instability starts to occur at $I_m = 1$ with a wavenumber $\omega_c^2 = 2I - \Delta = 2 - \Delta$. $\omega_c^{(2)}$ is always different from 0, giving rise to Turing instability.

In the bistable case, $\Delta \geq \sqrt{3}$: The system becomes unstable under the following conditions:

$$\begin{cases} \omega_c^{(1)} \text{ unstable} \quad I_- \leq I \leq I_+ \quad \Delta \geq \sqrt{3} \\ \\ \omega_c^{(1)} \text{ unstable \& minimum} \quad I_- \leq I \leq \frac{\Delta}{2} \quad \Delta > 2 \end{cases} \qquad \text{with} \quad I_\pm = \frac{2\Delta \pm \sqrt{\Delta^2 - 3}}{3}$$

$$\omega_c^{(2)} \text{ unstable \& minimum} \begin{cases} \Delta < 2 \\ I \geq I_m \end{cases} \quad \text{or} \quad \begin{cases} \Delta \geq 2 \\ I \geq \frac{\Delta}{2} \end{cases}$$

A.4 Conclusion

In this appendix, we conducted a detailed Linear Stability Analysis (LSA) to examine the conditions under which the Homogeneous Steady State (HSS) of the Lugiato-Lefever Equation (LLE) becomes unstable. The primary objective was to investigate the onset of Modulation Instability (MI) and the formation of solitons, phenomena that play a crucial role in nonlinear optical systems. By deriving the dispersion relation and examining the eigenvalues, we have identified key parameters that dictate the stability of the system.

This analysis serves as a foundational tool for understanding the rich dynamics that can arise in Kerr frequency combs and similar systems. It provides a mathematical framework for predicting the conditions under which pattern formation, such as Turing structures, and soliton dynamics occur. Furthermore, the identification of critical thresholds for the onset of MI aids in distinguishing between different regimes of operation—monostable and bistable—highlighting the sensitivity of the system to changes in detuning and intracavity intensity.

These results, referenced in Chapter 1, underscore the significance of LSA in shaping the stability and dynamics of nonlinear optical systems and will serve as a theoretical basis for further exploration into more complex behaviors, including chaos and turbulence.

Appendix B

Multivariate Transfer Entropy of two coupled Rössler systems

Outline of the current chapter

This appendix provides the computational framework for applying Multivariate Transfer Entropy to two coupled Rössler systems. The primary objective is to test and validate the Multivariate Transfer Entropy methodology in a known chaotic system, which serves as a preliminary step toward its application in wind dynamics analysis. By studying the directional flow of information between these coupled chaotic systems, we aim to validate and refine our Transfer Entropy (TE) methodology, ensuring its robustness and applicability to more complex, real-world scenarios, such as the analysis of wind data.

B.1 Introduction

In Chapter 1, we introduced the field of dynamical systems and chaos theory, specifically discussing the Lorenz and Rössler systems as paradigms of chaotic behavior. We explored their sensitivity to initial conditions and the phenomenon popularly known as the "butterfly effect." Building on this foundational knowledge, we now focus on the interaction between two chaotic systems, specifically the Rössler attractor, introduced by Otto Rössler in 1976 [29].

The Rössler system, despite its relatively simple set of nonlinear differential equations, exhibits rich and complex behavior, making it a popular model for studying chaos. In Section 1.1.4, we provided a small introduction to the Rössler system, describing its equations and the chaotic dynamics that arise under specific parameter settings.

When two Rössler systems are coupled, they can influence each other's behavior, leading to a variety of complex phenomena such as synchronization, where the systems' trajectories become correlated, and phase locking, where their oscillations lock into a consistent phase relationship.

To quantitatively analyze the interactions and information flow between coupled chaotic systems, we use a measure called Multivariate Transfer Entropy. TE is an information-theoretic quantity that captures the directional flow of information between time series, making it well-suited for studying the dynamics of coupled systems. In the following sections, we describe the methodology for calculating Multivariate TE between two coupled Rössler systems, highlighting the insights it provides into their complex interactions.

B.2 Coupling of Two Rössler Systems

When two chaotic systems like the Rössler attractor are coupled, the resulting dynamics can be complex and intriguing. The coupled Rössler systems are described by the following set of differential equations:

$$\begin{cases} \dot{x}_1 = -\omega_1 y_1 - z_1, \\ \dot{y}_1 = \omega_1 x_1 + a y_1, \\ \dot{z}_1 = b + z_1(x_1 - c), \end{cases} \tag{B.1}$$

for the first system, and

$$\begin{cases} \dot{x}_2 = -\omega_2 y_2 - z_2 + \epsilon(x_1 - x_2), \\ \dot{y}_2 = \omega_2 x_2 + a y_2, \\ \dot{z}_2 = b + z_2(x_2 - c), \end{cases} \tag{B.2}$$

for the second system, where ω_1 and ω_2 are the natural frequencies of the first and second Rössler systems, respectively, and ϵ represents the coupling strength between the two systems. The coupling term $\epsilon(x_1 - x_2)$ in the equation for $\dot{x}_2$ introduces the interaction between the systems, with $\epsilon = 0$ corresponding to no coupling and larger values of ϵ corresponding to stronger coupling. In our case, the following parameters were used: the constants $a = 0.15$, $b = 0.2$, and $c = 10$ were chosen to ensure chaotic behavior within each individual Rössler system. The coupling strength ϵ was varied within the range $[0, 1]$. We worked with the frequency pair $\omega_1 = 0.5$ and $\omega_2 = 2.515$. The time parameters for the simulations were set with a time step of 0.01 and a time span from 0 to 200. The first 1000 data points were discarded, whereas the next 19000 ones were saved. These parameters ensure a sufficiently long period for the systems to exhibit their chaotic and coupled behaviors while maintaining computational efficiency.

The dynamics of the coupled Rössler systems depend heavily on the coupling strength ϵ. When ϵ is varied, the coupled Rössler systems can exhibit a range of behaviors. For weak coupling, the systems may behave independently, maintaining their individual chaotic trajectories. However, as the coupling strength increases, the systems begin to influence each other more significantly. At higher coupling strengths, the systems may synchronize, meaning their trajectories become correlated, and they exhibit similar dynamic behavior. This transition from independent to synchronized behavior illustrates how information and influence propagate between interacting chaotic systems.

Understanding the nature of this interaction requires a quantitative measure of directional influence, which is where Multivariate Transfer Entropy comes into play. TE quantifies the amount of information transferred from one system to another, providing insights into the directional dependencies within the coupled systems.

B.3 Methodology for Calculating Multivariate Transfer Entropy

To investigate the interactions and information flow between coupled Rössler systems, we employ Multivariate Transfer Entropy, an information-theoretic measure that extends the concept of TE to multivariate time series. This measure helps us understand how information about the state of one system can predict the future state of another system, providing insights into the directional dependencies within the coupled systems. TE from system Y to system X is defined as:

$$T_{Y\to X} = \sum p(x_{n+1}, x_n, y_n) \log \frac{p(x_{n+1}|x_n, y_n)}{p(x_{n+1}|x_n)}, \tag{B.3}$$

where x_n and y_n represent the states of systems X and Y at time n, respectively. The joint probability distribution $p(x_{n+1}, x_n, y_n)$ captures the likelihood of transitions in X conditioned on the state of Y.

To compute this measure, we first need to reconstruct the state space of each system using time-delay embedding, a technique that transforms the time series data into a multidimensional phase space [134, 200]. This phase space allows us to capture the system's dynamics by incorporating past states, which is crucial for understanding the interactions between coupled systems. The process of time-delay embedding, based on Takens' method [134], allows us to reconstruct the state space of a system from scalar time series data, capturing the system's dynamics in a multidimensional space. This technique is essential for analyzing the interactions and information flow between coupled Rössler systems.

The data for the two coupled Rössler systems, represented as multivariate time series X and Y, are used for various coupling strengths ϵ. Each system's state is represented by three variables (x, y, z). The time-delay embedding technique transforms these time series into higher-dimensional phase spaces, accurately capturing their dynamics.

For each time series, we need to determine the optimal embedding dimension d and time delay τ. The embedding dimension defines the number of past states to include in the reconstruction, while the time delay specifies the interval between these states. These parameters are critical for accurately capturing the system's dynamics and interactions. To determine these optimal embedding parameters, we employed the Java Information Dynamics Toolkit (JIDT). Specifically, we used mutual information to select the optimal time delays τ for each variable, and the False Nearest Neighbors (FNN) method to determine the appropriate embedding dimension d. Mutual information measures the amount of information shared between two variables, and by analyzing mutual information over different time lags, we can identify the delay that maximizes the information content. For the embedding dimension d, the False Nearest Neighbors (FNN) method identifies the appropriate dimension by analyzing how the proportion of false neighbors decreases as the dimensionality increases. A false neighbor occurs when points close in a lower-dimensional space are not close in a higher-dimensional space, indicating that the lower-dimensional representation is insufficient to capture the dynamics.

In our implementation, we used the `TransferEntropyCalculatorKraskov` class from JIDT, setting it to automatically determine the optimal embedding parameters using Ragwitz's criterion. This method performs a search for the best embedding dimension and delay by maximizing the predictive power of the past states.

Once the optimal embedding parameters are determined, we reconstruct the phase space for each variable in the time series. This involves creating phase space vectors of the form $\mathbf{W}_k = \left(x_k, x_{k+\tau}, \ldots, x_{k+(d-1)\tau}\right)$, for both X and Y, capturing the dynamics of each system in a higher-dimensional space. The reconstructed phase spaces are then used to calculate the Multivariate Transfer Entropy between the systems. The TE quantifies the directional information flow between the systems, providing insights into their interactions. For this calculation, we use the Kraskov-Stögbauer-Grassberger (KSG) estimator provided by the JIDT, which is a non-parametric method, well-suited for continuous variables and is based on nearest-neighbor statistics, making it robust for high-dimensional data.

By varying the coupling strength ϵ and calculating the TE for each value, we analyze how information flow between the systems changes with increasing interaction. This method for time-delay reconstruction of phase space and calculation of Multivariate Transfer Entropy provides a robust framework for analyzing the directional dependencies and information flow between coupled chaotic systems. By accurately reconstructing the phase space and using information-theoretic measures, we gain deep insights into the complex dynamics of coupled Rössler systems, paving the way for applying these techniques to more complex systems, such as wind data analysis. The application of mutual information for delay selection and the FNN method for embedding dimension determination ensures that our phase space reconstruction is both accurate and representative of the system's dynamics. The use of the KSG estimator for TE calculation allows us to capture the intricate information transfer between the coupled systems, providing a comprehensive understanding of their interactions.

B.4 Results and Analysis

In this section, we present the findings from our study on the Multivariate Transfer Entropy of two coupled Rössler systems. These results offer insights into the directional dependencies and information flow between the coupled chaotic systems.

B.4.1 Multivariate Transfer Entropy of Coupled Rössler Systems

The Multivariate Transfer Entropy from system X to system Y and vice versa is shown in Figure B.1. This figure illustrates the TE values in nats as a function of the coupling strength ϵ.

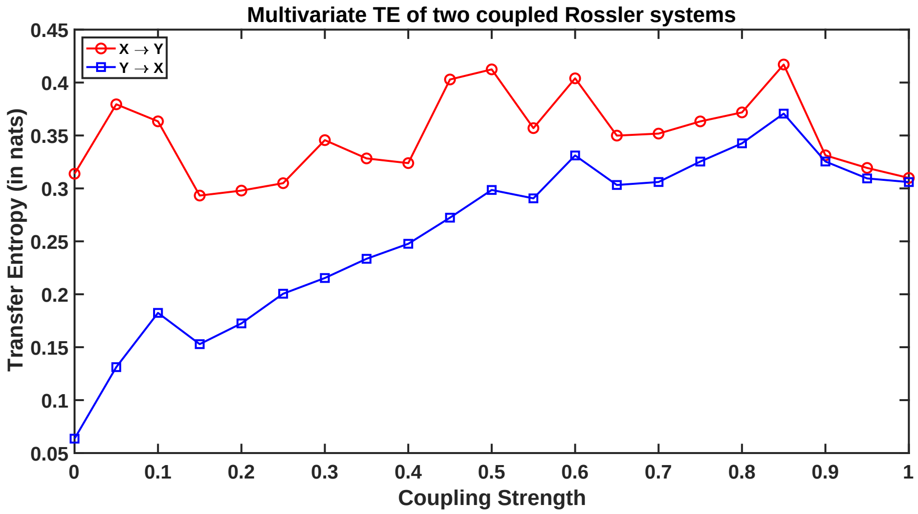


Figure B.1: Multivariate TE of two coupled Rössler systems. TE from $X \rightarrow Y$ is shown in red, while TE from $Y \rightarrow X$ is shown in blue.

The TE from $X \rightarrow Y$ (red curve) is generally higher than the TE from $Y \rightarrow X$ (blue curve) across most coupling strengths. This indicates a stronger influence of system X on system Y compared to the reverse. As the coupling strength increases, the TE from $Y \rightarrow X$ also increases, indicating a growing influence of Y on X.

Interestingly, at a coupling strength of $\epsilon = 1$, the TE values in both directions are nearly equal. This observation aligns with the suggestion made in the literature that synchronization takes place at a coupling of about 1 [201]. In a synchronized state, the systems influence each other equally, leading to nearly equal TE values in both directions. This indicates that at $\epsilon = 1$, the coupled Rössler systems are likely synchronized, exhibiting similar dynamic behavior and mutual influence. In another study by Krakovska et al. [202], the Conditional Mutual Information (CMI) between the two coupled Rössler systems was analyzed, showing a similar behavior to our findings. The CMI results also suggest that synchronization occurs at a coupling strength of about 1, supporting our observations.

The method used by Mao and Shang in their study of multivariate transfer entropy also focused on the time-delay reconstruction of phase space [203]. However, there are some differences between our approach and theirs. While both methods utilize time-delay embedding to reconstruct the phase space, their TE calculation differs in the way the embedding dimensions and time delays are determined. Mao and Shang employed mutual information to find the time delays for scalar time series, whereas we utilized the JIDT package to automate the determination of optimal embedding parameters using Ragwitz's criterion. Additionally, Mao and Shang showed that the TE in both directions decreases as the coupling strength increases, whereas our results indicate that the TE from $Y \to X$ increases with coupling strength, suggesting a growing influence of Y on X as the coupling becomes stronger. Notably, in both studies, the TE values from $X \to Y$ and $Y \to X$ become nearly equal when the coupling strength is close to 1, indicating synchronization between the systems. Despite these methodological differences, the overall trend and behavior of TE in response to varying coupling strengths remain consistent, highlighting the robustness of the TE approach in capturing directional dependencies in coupled chaotic systems.

B.4.2 Normalized Directionality Transfer Entropy (NDTE)

To further understand the directionality of the coupling, we computed the Normalized Directionality Transfer Entropy (NDTE) [203] defined as:

$$\mathrm{NDTE} = \frac{\mathrm{TE}_{Y\to X} - \mathrm{TE}_{X\to Y}}{\mathrm{TE}_{Y\to X} + \mathrm{TE}_{X\to Y}}. \tag{B.4}$$

The NDTE values range from -1 to 1, where positive values indicate that X is the driver, and negative values indicate that Y is the driver.

Figure B.2 shows the NDTE values as a function of the coupling strength ϵ. At $\epsilon = 0$, the NDTE is positive, indicating that system X drives system Y. As the coupling strength increases, the NDTE decreases, reflecting a reduction in the dominance of X over Y. At high coupling strengths ($\epsilon > 0.9$), the NDTE approaches zero, implying an almost symmetric bidirectional influence.

These findings indicate that at low coupling strengths, system X predominantly drives system Y. As the coupling strength increases, the influence of Y on X becomes more significant, leading to a more balanced interaction. This transition from unidirectional to bidirectional influence highlights the complex dynamics and information flow between the coupled Rössler systems.

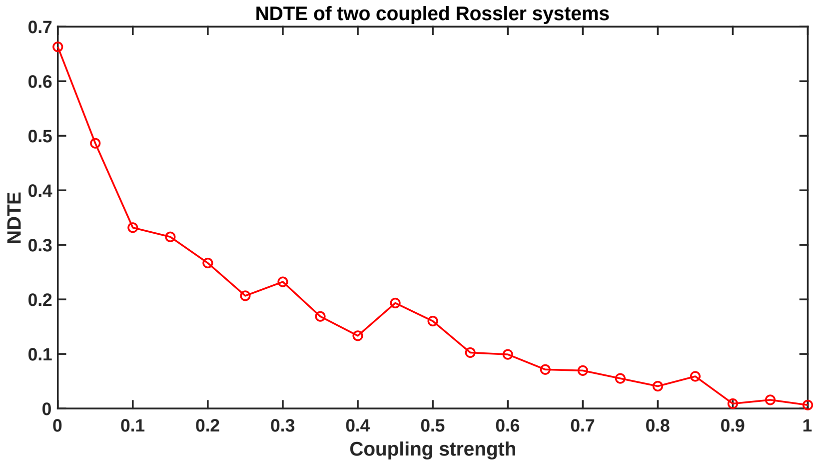


Figure B.2: Normalized Directionality Transfer Entropy (NDTE) of two coupled Rössler systems.

B.5 Conclusion

In this appendix, we developed a method to compute the Multivariate Transfer Entropy and applied it to study the interactions and information flow between two coupled Rössler systems. Our methodology involved phase-space reconstruction using time-delay embedding, determining optimal embedding parameters with the Java Information Dynamics Toolkit (JIDT) package, and employing the Kraskov-Stögbauer-Grassberger (KSG) estimator to calculate the TE. The results demonstrated the effectiveness of our approach, revealing key insights into the directional dependencies within the coupled chaotic systems.

Our findings are in accord with the existing literature, particularly the synchronization behavior observed at a coupling strength of approximately 1. This was evidenced by the nearly equal TE values in both directions, indicating a balanced mutual influence between the systems. The behavior of the Conditional Mutual Information (CMI) further supported these observations, aligning with previous studies that analyzed similar coupled chaotic systems.

Additionally, the computed Normalized Directionality Transfer Entropy (NDTE) offered further insights into the directionality of the coupling, revealing a transition from unidirectional to bidirectional influence as coupling strength increased.

This appendix validated our TE methodology for coupled Rössler systems, paving the way for its application in more complex, real-world scenarios. The methodology developed here will be employed in Chapter 4, where we will apply it to analyze wind dynamics. By doing so, we aim to uncover the directional information flow within wind data from multiple meteorological stations, testing the robustness of our method in a practical context and contributing to a deeper understanding and prediction of wind behavior, which is critical for atmospheric and environmental applications.

Bibliography

[1] E. N. Lorenz, “Deterministic Nonperiodic Flow”, Journal of the Atmospheric Sciences **20**, 130–141 (1963).

[2] J. M. Dudley, G. Genty, and B. J. Eggleton, “Harnessing and control of optical rogue waves in supercontinuum generation”, Optics Express **16**, 3644 (2008).

[3] A. Mussot, A. Kudlinski, M. Kolobov, E. Louvergneaux, M. Douay, and M. Taki, “Observation of extreme temporal events in CW-pumped supercontinuum”, Optics Express **17**, 17010 (2009).

[4] F. T. Arecchi, U. Bortolozzo, A. Montina, and S. Residori, “Granularity and Inhomogeneity Are the Joint Generators of Optical Rogue Waves”, Physical Review Letters **106**, 153901 (2011).

[5] G. Genty, L. Salmela, J. M. Dudley, D. Brunner, A. Kokhanovskiy, S. Kobtsev, and S. K. Turitsyn, “Machine learning and applications in ultrafast photonics”, Nature Photonics **15**, 91–101 (2021).

[6] R. I. Woodward and E. J. R. Kelleher, “Towards ‘smart lasers’: self-optimisation of an ultrafast pulse source using a genetic algorithm”, Scientific Reports **6**, 37616 (2016).

[7] T. Baumeister, S. L. Brunton, and J. Nathan Kutz, “Deep learning and model predictive control for self-tuning mode-locked lasers”, Journal of the Optical Society of America B **35**, 617 (2018).

[8] T. Zahavy, A. Dikopoltsev, D. Moss, G. I. Haham, O. Cohen, S. Mannor, and M. Segev, “Deep learning reconstruction of ultrashort pulses”, Optica **5**, 666 (2018).

[9] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, “Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach”, Physical review letters **120**, 024102 (2018).

[10] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, “Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data”, Chaos: An Interdisciplinary Journal of Nonlinear Science **27**, 121102 (2017).

[11]P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and P. Koumoutsakos, "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks", Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **474**, 20170844 (2018).

[12]S. T. Thornton and J. B. Marion, *Classical Dynamics of Particles and Systems* (Brooks/Cole, 2004).

[13]R. M. May, "Simple mathematical models with very complicated dynamics", Nature **261**, 459–467 (1976).

[14]*Philosophiae Naturalis Principia Mathematica. Auctore Js. Newton ...* (1687).

[15]H. (.-1. A. du texte Poincaré, *Les méthodes nouvelles de la mécanique céleste. Invariants intégraux ; solutions périodiques du deuxième genre ; solutions doublement asymptotiques / par H. Poincaré,...* (1892/1899).

[16]H. (.-1. A. du texte Poincaré, *La Science et l'hypothèse / par H. Poincaré,...* (1902).

[17]B. van der Pol, "The Nonlinear Theory of Electric Oscillations", Proceedings of the Institute of Radio Engineers **22**, 1051–1086 (1934).

[18]E. A. Jackson, *Perspectives of Nonlinear Dynamics*, 1st ed. (Cambridge University Press, Sept. 1989).

[19]H. Poincaré, *The Three-Body Problem and the Equations of Dynamics*, Vol. 443, Astrophysics and Space Science Library (Springer International Publishing, Cham, 2017).

[20]T.-Y. Li and J. A. Yorke, "Period Three Implies Chaos", The American Mathematical Monthly **82**, 985–992 (1975).

[21]T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology", Journal of Statistical Physics **65**, 579–616 (1991).

[22]J. P. Gollub and H. L. Swinney, "Onset of Turbulence in a Rotating Fluid", Physical Review Letters **35**, 927–930 (1975).

[23]T.-Y. Li and J. A. Yorke, "Period Three Implies Chaos", in *The Theory of Chaotic Attractors*, edited by B. R. Hunt, T.-Y. Li, J. A. Kennedy, and H. E. Nusse (Springer New York, New York, NY, 2004), pp. 77–84.

[24]E. Ott, C. Grebogi, and J. A. Yorke, "Controlling chaos", Physical Review Letters **64**, 1196–1199 (1990).

[25]É. Ghys, "The Lorenz Attractor, a Paradigm for Chaos", in *Chaos: Poincaré Seminar 2010*, edited by B. Duplantier, S. Nonnenmacher, and V. Rivasseau (Springer, Basel, 2013), pp. 1–54.

[26]S. Bouali, *A 3D Strange Attractor with a Distinctive Silhouette. The Butterfly Effect Revisited*, Mar. 2014.

[27] L. Prandtl, "Uber Flussigkeitsbewegung bei sehr kleiner Reibung", Verhandl. 3rd Int. Math. Kongr. Heidelberg (1904), Leipzig (1905).

[28] L. Rayleigh, "On the stability, or instability, of certain fluid motions", Proceedings of the London Mathematical Society **1**, 57–72 (1879).

[29] O. Rössler, "An equation for continuous chaos", Physics Letters A **57**, 397–398 (1976).

[30] A. Scott, ed., *Encyclopedia of Nonlinear Science*, 0th ed. (Routledge, May 2006).

[31] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, "Reservoir observers: Model-free inference of unmeasured variables in chaotic systems", Chaos: An Interdisciplinary Journal of Nonlinear Science **27**, 041102 (2017).

[32] E. Ott, *Chaos in Dynamical Systems*, 2nd ed. (Cambridge University Press, Aug. 2002).

[33] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, 2018).

[34] F. Takens, "Detecting strange attractors in turbulence", in *Dynamical Systems and Turbulence, Warwick 1980*, Vol. 898, edited by D. Rand and L.-S. Young (Springer Berlin Heidelberg, Berlin, Heidelberg, 1981), pp. 366–381.

[35] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series", Physica D: Nonlinear Phenomena **16**, 285–317 (1985).

[36] M. C. Cross and P. C. Hohenberg, "Pattern formation outside of equilibrium", Reviews of Modern Physics **65**, 851–1112 (1993).

[37] Y. Kuramoto, "Chemical Turbulence", in *Chemical Oscillations, Waves, and Turbulence*, Vol. 19 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1984), pp. 111–140.

[38] I. Prigogine and G. Nicolis, "Self-Organisation in Nonequilibrium Systems: Towards A Dynamics of Complexity", in *Bifurcation Analysis*, edited by M. Hazewinkel, R. Jurkovich, and J. H. P. Paelinck (Springer Netherlands, Dordrecht, 1985), pp. 3–12.

[39] I. S. Aranson and L. Kramer, "The world of the complex Ginzburg-Landau equation", Reviews of Modern Physics **74**, 99–143 (2002).

[40] Y. Kuramoto, "Diffusion-Induced Chaos in Reaction Systems", Progress of Theoretical Physics Supplement **64**, 346–367 (1978).

[41] T. Udem, R. Holzwarth, and T. W. Hänsch, "Optical frequency metrology", Nature **416**, 233–237 (2002).

[42] I. Coddington, W. C. Swann, L. Nenadovic, and N. R. Newbury, "Rapid and precise absolute distance measurements at long range", Nature Photonics **3**, 351–356 (2009).

[43] P. Trocha, M. Karpov, D. Ganin, M. H. P. Pfeiffer, A. Kordts, S. Wolf, J. Krockenberger, P. Marin-Palomo, C. Weimann, S. Randel, W. Freude, T. J. Kippenberg, and C. Koos, "Ultrafast optical ranging using microresonator soliton frequency combs", Science **359**, 887–891 (2018).

[44] E. Obrzud, M. Rainer, A. Harutyunyan, M. H. Anderson, J. Liu, M. Geiselmann, B. Chazelas, S. Kundermann, S. Lecomte, M. Cecconi, A. Ghedina, E. Molinari, F. Pepe, F. Wildi, F. Bouchy, T. J. Kippenberg, and T. Herr, "A microphotonic astrocomb", Nature Photonics **13**, 31–35 (2019).

[45] T. Steinmetz, T. Wilken, C. Araujo-Hauck, R. Holzwarth, T. W. Hänsch, L. Pasquini, A. Manescau, S. D'Odorico, M. T. Murphy, T. Kentischer, W. Schmidt, and T. Udem, "Laser Frequency Combs for Astronomical Observations", Science **321**, 1335–1337 (2008).

[46] I. Coddington, N. Newbury, and W. Swann, "Dual-comb spectroscopy", Optica **3**, 414–426 (2016).

[47] T. W. Hänsch, "Nobel Lecture: Passion for precision", Reviews of Modern Physics **78**, 1297–1309 (2006).

[48] H. Timmers, A. Kowligy, A. Lind, F. C. Cruz, N. Nader, M. Silfies, G. Ycas, T. K. Allison, P. G. Schunemann, S. B. Papp, and S. A. Diddams, "Molecular fingerprinting with bright, broadband infrared frequency combs", Optica **5**, 727–732 (2018).

[49] L. A. Lugiato, F. Prati, M. L. Gorodetsky, and T. J. Kippenberg, "From the Lugiato–Lefever equation to microresonator-based soliton Kerr frequency combs", Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **376**, 20180113 (2018).

[50] D. J. Jones, S. A. Diddams, J. K. Ranka, A. Stentz, R. S. Windeler, J. L. Hall, and S. T. Cundiff, "Carrier-Envelope Phase Control of Femtosecond Mode-Locked Lasers and Direct Optical Frequency Synthesis", Science **288**, 635–639 (2000).

[51] E. Myslivets, B. P. P. Kuo, N. Alic, and S. Radic, "Generation of wideband frequency combs by continuous-wave seeding of multistage mixers with synthesized dispersion", Optics Express **20**, 3331–3344 (2012).

[52] A. Hugi, G. Villares, S. Blaser, H. C. Liu, and J. Faist, "Mid-infrared frequency comb based on a quantum cascade laser", Nature **492**, 229–233 (2012).

[53]M. Zimmermann, C. Gohle, R. Holzwarth, T. Udem, and T. W. Hänsch, "Optical clockwork with an offset-free difference-frequency comb: accuracy of sum- and difference-frequency generation", Optics Letters **29**, 310–312 (2004).

[54]P. Del'Haye, A. Schliesser, O. Arcizet, T. Wilken, R. Holzwarth, and T. J. Kippenberg, "Optical frequency comb generation from a monolithic microresonator", Nature **450**, 1214–1217 (2007).

[55]K. J. Vahala, "Optical microcavities", Nature **424**, 839–846 (2003).

[56]T. J. Kippenberg, S. M. Spillane, and K. J. Vahala, "Kerr-Nonlinearity Optical Parametric Oscillation in an Ultrahigh- Q Toroid Microcavity", Physical Review Letters **93**, 083904 (2004).

[57]A. A. Savchenkov, A. B. Matsko, D. Strekalov, M. Mohageg, V. S. Ilchenko, and L. Maleki, "Low Threshold Optical Oscillations in a Whispering Gallery Mode C a F 2 Resonator", Physical Review Letters **93**, 243905 (2004).

[58]L. A. Lugiato and R. Lefever, "Spatial Dissipative Structures in Passive Optical Systems", Physical Review Letters **58**, 2209–2211 (1987).

[59]S. Coen and M. Erkintalo, "Universal scaling laws of Kerr frequency combs", Optics Letters **38**, 1790 (2013).

[60]F. Leo, S. Coen, P. Kockaert, S.-P. Gorza, P. Emplit, and M. Haelterman, "Temporal cavity solitons in one-dimensional Kerr media as bits in an all-optical buffer", Nature Photonics **4**, 471–476 (2010).

[61]S. Coen and M. Haelterman, "Modulational Instability Induced by Cavity Boundary Conditions in a Normally Dispersive Optical Fiber", Physical Review Letters **79**, 4139–4142 (1997).

[62]M. Haelterman, S. Trillo, and S. Wabnitz, "Dissipative modulation instability in a nonlinear dispersive ring cavity", Optics Communications **91**, 401–407 (1992).

[63]K. Ikeda, "Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system", Optics Communications **30**, 257–261 (1979).

[64]A. Montina, U. Bortolozzo, S. Residori, and F. T. Arecchi, "Non-Gaussian Statistics and Extreme Waves in a Nonlinear Optical Cavity", Physical Review Letters **103**, 173901 (2009).

[65]K. Nozaki and N. Bekki, "Solitons as attractors of a forced dissipative nonlinear Schrödinger equation", Physics Letters A **102**, 383–386 (1984).

[66]D. J. Kaup and A. C. Newell, "Theory of nonlinear oscillating dipolar excitations in one-dimensional condensates", Physical Review B **18**, 5162–5167 (1978).

[67]R. Kapral and K. Showalter, *Chemical Waves and Patterns* (Springer Science & Business Media, Dec. 2012).

[68]"The chemical basis of morphogenesis", Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **237**, 37–72 (1952).

[69]V. Castets, E. Dulos, J. Boissonade, and P. De Kepper, "Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern", Physical Review Letters **64**, 2953–2956 (1990).

[70]K. Tai, A. Hasegawa, and A. Tomita, "Observation of modulational instability in optical fibers", Physical Review Letters **56**, 135–138 (1986).

[71]M. Nakazawa, K. Suzuki, and H. A. Haus, "Modulational instability oscillation in nonlinear dispersive ring cavity", Physical Review A **38**, 5193–5196 (1988).

[72]T. J. Kippenberg, A. L. Gaeta, M. Lipson, and M. L. Gorodetsky, "Dissipative Kerr solitons in optical microresonators", Science **361**, eaan8083 (2018).

[73]A. Scroggie, W. Firth, G. McDonald, M. Tlidi, R. Lefever, and L. Lugiato, "Pattern formation in a passive Kerr cavity", Chaos, Solitons & Fractals **4**, 1323–1354 (1994).

[74]S. Coen and M. Haelterman, "Competition between modulational instability and switching in optical bistability", Optics Letters **24**, 80 (1999).

[75]M. Tlidi, P. Mandel, and R. Lefever, "Localized structures and localized patterns in optical bistability", Physical Review Letters **73**, 640–643 (1994).

[76]W. J. Firth and A. J. Scroggie, "Optical Bullet Holes: Robust Controllable Localized States of a Nonlinear Cavity", Physical Review Letters **76**, 1623–1626 (1996).

[77]P. Coullet, C. Riera, and C. Tresser, "Stable Static Localized Structures in One Dimension", Physical Review Letters **84**, 3069–3072 (2000).

[78]L. Lugiato, "Introduction to the feature section on cavity solitons: An overview", IEEE Journal of Quantum Electronics **39**, 193–196 (2003).

[79]F. Leo, L. Gelens, P. Emplit, M. Haelterman, and S. Coen, "Dynamics of one-dimensional Kerr cavity solitons", Optics Express **21**, 9180 (2013).

[80]P. Parra-Rivas, D. Gomila, F. Leo, S. Coen, and L. Gelens, "Third-order chromatic dispersion stabilizes Kerr frequency combs", Optics Letters **39**, 2971 (2014).

[81]K. Nozaki and N. Bekki, "Chaotic Solitons in a Plasma Driven by an rf Field", Journal of the Physical Society of Japan **54**, 2363–2366 (1985).

[82]Y. K. Chembo, D. V. Strekalov, and N. Yu, “Spectrum and Dynamics of Optical Frequency Combs Generated with Monolithic Whispering Gallery Mode Resonators”, Physical Review Letters **104**, 103902 (2010).

[83]A. B. Matsko, W. Liang, A. A. Savchenkov, and L. Maleki, “Chaotic dynamics of frequency combs generated with continuously pumped nonlinear microresonators”, Optics Letters **38**, 525 (2013).

[84]A. Coillet and Y. K. Chembo, “Routes to spatiotemporal chaos in Kerr optical frequency combs”, Chaos: An Interdisciplinary Journal of Nonlinear Science **24**, 013113 (2014).

[85]A. Coillet, J. Dudley, G. Genty, L. Larger, and Y. K. Chembo, “Optical rogue waves in whispering-gallery-mode resonators”, Physical Review A **89**, 013835 (2014).

[86]V. Ruban, Y. Kodama, M. Ruderman, J. Dudley, R. Grimshaw, P. McClintock, M. Onorato, C. Kharif, E. Pelinovsky, T. Soomere, G. Lindgren, N. Akhmediev, A. Slunyaev, D. Solli, C. Ropers, B. Jalali, F. Dias, and A. Osborne, “Rogue waves – towards a unifying concept?: Discussions and debates”, The European Physical Journal Special Topics **185**, 5–15 (2010).

[87]N. Akhmediev, J. M. Dudley, D. R. Solli, and S. K. Turitsyn, “Recent progress in investigating optical rogue waves”, Journal of Optics **15**, 060201 (2013).

[88]M. Anderson, F. Leo, S. Coen, M. Erkintalo, and S. G. Murdoch, “Observations of spatiotemporal instabilities of temporal cavity solitons”, Optica **3**, 1071 (2016).

[89]F. Mitschke, G. Steinmeyer, and A. Schwache, “Generation of one-dimensional optical turbulence”, Physica D: Nonlinear Phenomena **96**, 251–258 (1996).

[90]G. Steinmeyer, A. Schwache, and F. Mitschke, “Quantitative characterization of turbulence in an optical experiment”, Physical Review E **53**, 5399–5402 (1996).

[91]P. Manneville, “Dissipative structures and weak turbulence”, in *Chaos — The Interplay Between Stochastic and Deterministic Behaviour*, Vol. 457, edited by P. Garbaczewski, M. Wolf, and A. Weron (Springer Berlin Heidelberg, Berlin, Heidelberg, 1995), pp. 257–272.

[92]D. Ruelle, “Large volume limit of the distribution of characteristic exponents in turbulence”, Communications in Mathematical Physics **87**, 287–302 (1982).

[93]A. Pikovsky and A. Politi, *Lyapunov Exponents: A Tool to Explore Complex Dynamics*, 1st ed. (Cambridge University Press, Dec. 2015).

[94]G. Nicolis, *Introduction to Nonlinear Science*, 1st ed. (Cambridge University Press, June 1995).

[95]M. G. Clerc and N. Verschueren, “Quasiperiodicity route to spatiotemporal chaos in one-dimensional pattern-forming systems”, Physical Review E **88**, 052916 (2013).

[96]F. Selmi, S. Coulibaly, Z. Loghmari, I. Sagnes, G. Beaudoin, M. G. Clerc, and S. Barbay, “Spatiotemporal Chaos Induces Extreme Events in an Extended Microcavity Laser”, Physical Review Letters **116**, 013901 (2016).

[97]S. Coulibaly, M. G. Clerc, F. Selmi, and S. Barbay, “Extreme events following bifurcation to spatiotemporal chaos in a spatially extended microcavity laser”, Physical Review A **95**, 023816 (2017).

[98]U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, 1995).

[99]C. Skokos, “The Lyapunov Characteristic Exponents and Their Computation”, in *Dynamics of Small Solar System Bodies and Exoplanets*, Vol. 790, edited by J. J. Souchay and R. Dvorak (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), pp. 63–135.

[100]J. L. Kaplan and J. A. Yorke, “Chaotic behavior of multidimensional difference equations”, in *Functional Differential Equations and Approximation of Fixed Points*, Vol. 730, edited by H.-O. Peitgen and H.-O. Walther (Springer Berlin Heidelberg, Berlin, Heidelberg, 1979), pp. 204–227.

[101]S. Coulibaly, M. Taki, A. Bendahmane, G. Millot, B. Kibler, and M. G. Clerc, “Turbulence-Induced Rogue Waves in Kerr Resonators”, Physical Review X **9**, 011054 (2019).

[102]Z. Liu, M. Ouali, S. Coulibaly, M. G. Clerc, M. Taki, and M. Tlidi, “Characterization of spatiotemporal chaos in a Kerr optical frequency comb and in all fiber cavities”, Optics letters **42**, 1063–1066 (2017).

[103]S. Coulibaly, F. Bessin, M. G. Clerc, and A. Mussot, “Precursors-driven machine learning prediction of chaotic extreme pulses in Kerr resonators”, Chaos, Solitons & Fractals **160**, 112199 (2022).

[104]H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Institute for Nonlinear Science (Springer New York, New York, NY, 1996).

[105]D. R. Solli, C. Ropers, P. Koonath, and B. Jalali, “Optical rogue waves”, Nature **450**, 1054–1057 (2007).

[106]B. Kibler, J. Fatome, C. Finot, G. Millot, F. Dias, G. Genty, N. Akhmediev, and J. M. Dudley, “The Peregrine soliton in nonlinear fibre optics”, Nature Physics **6**, 790–795 (2010).

[107]H. L. D. de S. Cavalcante, M. Oriá, D. Sornette, E. Ott, and D. J. Gauthier, "Predictability and Suppression of Extreme Events in a Chaotic System", Physical Review Letters **111**, 198701 (2013).

[108]M. Onorato, A. R. Osborne, M. Serio, and S. Bertone, "Freak Waves in Random Oceanic Sea States", Physical Review Letters **86**, 5831–5834 (2001).

[109]H. S. Hudson, "Solar flares, microflares, nanoflares, and coronal heating", Solar Physics **133**, 357–369 (1991).

[110]L. STENFLO and P. K. SHUKLA, "Nonlinear acoustic–gravity waves", Journal of Plasma Physics **75**, 841–847 (2009).

[111]*On the Occurence of Extreme Events in Long-term Correlated and Multifractal Data Sets | SpringerLink*, https://link.springer.com/chapter/10.1007/978-3-7643-8907-9_11.

[112]S. E. Otto and C. W. Rowley, "Koopman Operators for Estimation and Control of Dynamical Systems", Annual Review of Control, Robotics, and Autonomous Systems **4**, 59–87 (2021).

[113]A. Asch, E. J. Brady, H. Gallardo, J. Hood, B. Chu, and M. Farazmand, "Model-assisted deep learning of rare extreme events from partial observations", Chaos: An Interdisciplinary Journal of Nonlinear Science **32**, 043112 (2022).

[114]S. Mukherjee, E. Osuna, and F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines", in Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop (Sept. 1997), pp. 511–520.

[115]M. Ghil, M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, and F. Varadi, "Advanced spectral methods for climatic time series", Reviews of geophysics **40**, 3-1-3–41 (2002).

[116]E. Racah, C. Beckham, T. Maharaj, S. E. Kahou, Prabhat, and C. Pal, *ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events*, Nov. 2017.

[117]V. A. Pammi, M. G. Clerc, S. Coulibaly, and S. Barbay, "Extreme Events Prediction from Nonlocal Partial Information in a Spatiotemporally Chaotic Microcavity Laser", Physical Review Letters **130**, 223801 (2023).

[118]M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects", Science **349**, 255–260 (2015).

[119]C. L. Chen, A. Mahjoubfar, L.-C. Tai, I. K. Blaby, A. Huang, K. R. Niazi, and B. Jalali, "Deep Learning in Label-free Cell Classification", Scientific Reports **6**, 21471 (2016).

[120]M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, "Deep-learning-based ghost imaging", Scientific Reports **7**, 17865 (2017).

[121]C. F. Higham, R. Murray-Smith, M. J. Padgett, and M. P. Edgar, "Deep learning for real-time single-pixel video", Scientific Reports **8**, 2369 (2018).

[122]Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks", Light: Science & Applications **7**, 17141–17141 (2017).

[123]E. Giacoumidis, J. Wei, I. Aldaya, and L. P. Barry, *Exceeding the Nonlinear Shannon-Limit in Coherent Optical Communications by MIMO Machine Learning*, 2018.

[124]A. Lapedes and R. Farber, "Nonlinear signal processing using neural networks: Prediction and system modelling", (1987).

[125]K. Krischer, R. Rico-Martínez, I. G. Kevrekidis, H. H. Rotermund, G. Ertl, and J. L. Hudson, "Model identification of a spatiotemporally varying catalytic reaction", AIChE Journal **39**, 89–98 (1993).

[126]Liangyue Cao, Yiguang Hong, Haiping Fang, and Guowei He, "Predicting chaotic time series with wavelet networks", Physica D: Nonlinear Phenomena **85**, 225–238 (1995).

[127]S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", Neural Computation **9**, 1735–1780 (1997).

[128]A. Graves and N. Jaitly, "Towards End-To-End Speech Recognition with Recurrent Neural Networks", in Proceedings of the 31st International Conference on Machine Learning (June 2014), pp. 1764–1772.

[129]H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling", (2014).

[130]I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks", Advances in neural information processing systems **27** (2014).

[131]K. Cho, B. Van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation", in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014), pp. 1724–1734.

[132]D. Dong, H. Wu, W. He, D. Yu, and H. Wang, “Multi-Task Learning for Multiple Language Translation”, in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (2015), pp. 1723–1732.

[133]Z. C. Lipton, J. Berkowitz, and C. Elkan, *A Critical Review of Recurrent Neural Networks for Sequence Learning*, 2015.

[134]F. Takens, “Detecting strange attractors in turbulence”, in Dynamical Systems and Turbulence, Warwick 1980, edited by D. Rand and L.-S. Young (1981), pp. 366–381.

[135]E. Bradley and H. Kantz, “Nonlinear time-series analysis revisited”, Chaos: An Interdisciplinary Journal of Nonlinear Science **25**, 097610 (2015).

[136]B. O. Koopman, “Hamiltonian Systems and Transformation in Hilbert Space”, Proceedings of the National Academy of Sciences **17**, 315–318 (1931).

[137]B. Lusch, J. N. Kutz, and S. L. Brunton, “Deep learning for universal linear embeddings of nonlinear dynamics”, Nature Communications **9**, 4950 (2018).

[138]Z. Lu, B. R. Hunt, and E. Ott, “Attractor reconstruction by machine learning”, Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 061104 (2018).

[139]Z. Y. Wan, P. Vlachas, P. Koumoutsakos, and T. Sapsis, “Data-assisted reduced-order modeling of extreme events in complex dynamical systems”, PLOS ONE **13**, edited by D. Durstewitz, e0197704 (2018).

[140]Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult”, IEEE Transactions on Neural Networks **5**, 157–166 (1994).

[141]I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning* (MIT press, 2016).

[142]S. Hochreiter, “The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **06**, 107–116 (1998).

[143]R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks”, in International conference on machine learning (2013), pp. 1310–1318.

[144]F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, “An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting”, 10.48550/ARXIV.1705.04378 (2017).

[145]N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, “Time-series extreme event forecasting with neural networks at uber”, in International conference on machine learning, Vol. 34 (2017), pp. 1–5.

[146]Z. Y. Wan and T. P. Sapsis, "Machine learning the kinematics of spherical particles in fluid flows", Journal of Fluid Mechanics **857**, R2 (2018).

[147]G. Neofotistos, M. Mattheakis, G. D. Barmparis, J. Hizanidis, G. P. Tsironis, and E. Kaxiras, "Machine Learning With Observers Predicts Complex Spatiotemporal Behavior", Frontiers in Physics **7**, 24 (2019).

[148]K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, *On the Properties of Neural Machine Translation: Encoder-Decoder Approaches*, 2014.

[149]P. Werbos, "Backpropagation through time: what it does and how to do it", Proceedings of the IEEE **78**, 1550–1560 (Oct./1990).

[150]C. E. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal **27**, 379–423 (1948).

[151]P. Billingsley, *Ergodic theory and information*, Vol. 1 (Wiley New York, 1965).

[152]T. M. Cover, *Elements of information theory* (John Wiley & Sons, 1999).

[153]A. Kraskov, H. Stögbauer, and P. Grassberger, "Erratum: Estimating mutual information [Phys. Rev. E **69** , 066138 (2004)]", Physical Review E **83**, 019903 (2011).

[154]K. Hlavackovaschindler, M. Palus, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis", Physics Reports **441**, 1–46 (2007).

[155]T. Schreiber, "Measuring Information Transfer", Physical Review Letters **85**, 461–464 (2000).

[156]M. Wibral, N. Pampu, V. Priesemann, F. Siebenhühner, H. Seiwert, M. Lindner, J. T. Lizier, and R. Vicente, "Measuring Information-Transfer Delays", PLoS ONE **8**, edited by S. Hayasaka, e55809 (2013).

[157]N. Ay and D. Polani, "INFORMATION FLOWS IN CAUSAL NETWORKS", Advances in Complex Systems **11**, 17–41 (2008).

[158]J. T. Lizier, "JIDT: An Information-Theoretic Toolkit for Studying the Dynamics of Complex Systems", Frontiers in Robotics and AI **1**, 10.3389/frobt.2014.00011 (2014).

[159]S. Frenzel and B. Pompe, "Partial Mutual Information for Coupling Analysis of Multivariate Time Series", Physical Review Letters **99**, 204101 (2007).

[160]G. Gómez-Herrero, W. Wu, K. Rutanen, M. Soriano, G. Pipa, and R. Vicente, "Assessing Coupling Dynamics from an Ensemble of Time Series", Entropy **17**, 1958–1970 (2015).

[161] A. Kraskov, *Synchronization and interdependence measures and their applications to the electroencephalogram of epilepsy patients and clustering of data* (NIC-Secretariat, Research Centre Jülich, 2004).

[162] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", Econometrica **37**, 424 (1969).

[163] L. M. Zhao, D. Y. Tang, and A. Q. Liu, "Chaotic dynamics of a passively mode-locked soliton fiber ring laser", Chaos: An Interdisciplinary Journal of Nonlinear Science **16**, 013128 (2006).

[164] E. G. Turitsyna, S. V. Smirnov, S. Sugavanam, N. Tarasov, X. Shu, S. A. Babin, E. V. Podivilov, D. V. Churkin, G. Falkovich, and S. K. Turitsyn, "The laminar–turbulent transition in a fibre laser", Nature Photonics **7**, 783–786 (2013).

[165] S. Wabnitz, "Optical turbulence in fiber lasers", Optics Letters **39**, 1362 (2014).

[166] K. Ikeda, H. Daido, and O. Akimoto, "Optical Turbulence: Chaotic Behavior of Transmitted Light from a Ring Cavity", Physical Review Letters **45**, 709–712 (1980).

[167] K. Ikeda and O. Akimoto, "Instability Leading to Periodic and Chaotic Self-Pulsations in a Bistable Optical Cavity", Physical Review Letters **48**, 617–620 (1982).

[168] P. Grelu, *Nonlinear optical cavity dynamics: from microresonators to fiber lasers* (John Wiley & Sons, 2015).

[169] P. Parra-Rivas, D. Gomila, M. A. Matías, S. Coen, and L. Gelens, "Dynamics of localized and patterned structures in the Lugiato-Lefever equation determine the stability and shape of optical frequency combs", Physical Review A **89**, 043813 (2014).

[170] P. Parra-Rivas, D. Gomila, L. Gelens, and E. Knobloch, "Bifurcation structure of periodic patterns in the Lugiato-Lefever equation with anomalous dispersion", Physical Review E **98**, 042212 (2018).

[171] T. B. Benjamin and J. E. Feir, "The disintegration of wave trains on deep water Part 1. Theory", Journal of Fluid Mechanics **27**, 417–430 (1967).

[172] A. Pasquazi, M. Peccianti, L. Razzari, D. J. Moss, S. Coen, M. Erkintalo, Y. K. Chembo, T. Hansson, S. Wabnitz, P. Del'Haye, X. Xue, A. M. Weiner, and R. Morandotti, "Micro-combs: A novel generation of optical sources", Physics Reports, Micro-Combs: A Novel Generation of Optical Sources **729**, 1–81 (2018).

[173] M. C. Cross and P. C. Hohenberg, "Spatiotemporal Chaos", Science **263**, 1569–1570 (1994).

[174] S. Ciliberto and P. Bigazzi, “Spatiotemporal Intermittency in Rayleigh-Bénard Convection”, Physical Review Letters **60**, 286–289 (1988).

[175] F. Chollet, “Keras: Deep learning library for theano and tensorflow”, URL: https://keras. io/k **7**, T1 (2015).

[176] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, “{TensorFlow}: A System for {Large-Scale} Machine Learning”, in 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (2016), pp. 265–283.

[177] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014.

[178] W. Cousins and T. P. Sapsis, “Reduced-order precursors of rare events in unidirectional nonlinear water waves”, Journal of Fluid Mechanics **790**, 368–388 (2016).

[179] S. Hallerberg, E. G. Altmann, D. Holstein, and H. Kantz, “Precursors of extreme increments”, Physical Review E **75**, 016706 (2007).

[180] D. A. Egolf and H. S. Greenside, “Relation between fractal dimension and spatial correlation length for extensive chaos”, Nature **369**, 129–131 (1994).

[181] C. S. O’Hern, D. A. Egolf, and H. S. Greenside, “Lyapunov spectral analysis of a nonequilibrium Ising-like transition”, Physical Review E **53**, 3374–3386 (1996).

[182] F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, 1965).

[183] D. A. Egolf and H. S. Greenside, “Characterization of the Transition from Defect to Phase Turbulence”, Physical Review Letters **74**, 1751–1754 (1995).

[184] B. E. Flores, “A pragmatic view of accuracy measurement in forecasting”, Omega **14**, 93–98 (1986).

[185] S. K. Morley, T. V. Brito, and D. T. Welling, “Measures of Model Performance Based On the Log Accuracy Ratio”, Space Weather **16**, 69–88 (2018).

[186] S. Morley, *Alternatives to accuracy and bias metrics based on percentage errors for radiation belt modeling applications*, tech. rep. LA-UR–16-24592, 1260362 (July 2016), LA-UR–16-24592, 1260362.

[187] P. Aguilera-Rojas, M. Clerc, S. Echeverría-Alar, Y. Soupart, and M. Tlidi, “Fingerprint pattern bi-turbulence in a driven dissipative optical system”, Chaos, Solitons & Fractals **182**, 114851 (2024).

[188] V. V. Volkov and Y. Zhu, “Deterministic phase unwrapping in the presence of noise”, Optics letters **28**, 2156–2158 (2003).

[189]H. Rodriguez, J. J. Flores, L. A. Morales, C. Lara, A. Guerra, and G. Manjarrez, “Forecasting from incomplete and chaotic wind speed data”, Soft Computing **23**, 10119–10127 (2019).

[190]G. T. Bitsuamlak, T. Stathopoulos, and C. Bédard, “Numerical Evaluation of Wind Flow over Complex Terrain: Review”, Journal of Aerospace Engineering **17**, 135–145 (2004).

[191]R. G. Kavasseri and K. Seetharaman, “Day-ahead wind speed forecasting using f-ARIMA models”, Renewable Energy **34**, 1388–1393 (2009).

[192]M. Bilgili and B. Sahin, “Comparative analysis of regression and artificial neural network models for wind speed prediction”, Meteorology and Atmospheric Physics **109**, 61–72 (2010).

[193]S. Barhmi, O. Elfatni, and I. Belhaj, “Forecasting of wind speed using multiple linear regression and artificial neural networks”, Energy Systems **11**, 935–946 (2020).

[194]M. Bilgili, B. Sahin, and A. Yasar, “Application of artificial neural networks for the wind speed prediction of target station using reference stations data”, Renewable Energy **32**, 2350–2360 (2007).

[195]D. Bechrakis and P. Sparis, “Correlation of Wind Speed Between Neighboring Measuring Stations”, IEEE Transactions on Energy Conversion **19**, 400–406 (2004).

[196]M. Ragwitz and H. Kantz, “Markov models from data by simple nonlinear time series predictors in delay embedding spaces”, Physical Review E **65**, 056201 (2002).

[197]“Signal Processing Applications”, in *Introduction to signal processing* (Prentice-Hall, Inc., USA, Oct. 1995), p. 813.

[198]B. Chen, J. Li, and R. Ding, “Nonlinear local Lyapunov exponent and atmospheric predictability research”, Science in China Series D: Earth Sciences **49**, 1111–1120 (2006).

[199]J. Li and R. Ding, “Temporal–Spatial Distribution of Atmospheric Predictability Limit by Local Dynamical Analogs”, Monthly Weather Review **139**, 3265–3283 (2011).

[200]L. Cao, A. Mees, and K. Judd, “Dynamics from multivariate time series”, Physica D: Nonlinear Phenomena **121**, 75–88 (1998).

[201]A. Krakovská, J. Jakubík, M. Chvosteková, D. Coufal, N. Jajcay, and M. Paluš, “Comparison of six methods for the detection of causality in a bivariate time series”, Physical Review E **97**, 042207 (2018).

[202]A. Krakovská, J. Jakubík, H. Budáčová, and M. Holecyová, *Causality studied in reconstructed state space. Examples of uni-directionally connected chaotic systems*, 2015.

[203]X. Mao and P. Shang, "Transfer entropy between multivariate time series", Communications in Nonlinear Science and Numerical Simulation **47**, 338–347 (2017).

Machine Learning-Assisted Spatiotemporal Chaos Forecasting

Abstract

From towering rogue waves to powerful winds, extreme events can disrupt natural systems and human activity without warning. Though seemingly unpredictable, these events often arise from the complex dynamics of chaotic systems, particularly spatiotemporal chaos, where patterns unfold across both time and space. In this thesis, we study extreme events in optical systems, focusing on an optical fiber ring resonator modeled by the Lugiato-Lefever equation. This setup provides a controlled environment to analyze the chaotic behaviors that lead to such phenomena. Recent advancements in machine learning, especially neural networks, offer new tools for predicting chaotic dynamics. However, long-term forecasting remains challenging due to chaos's inherent unpredictability. We propose extending the prediction horizon using information theory methods, like transfer entropy, to identify local regions contributing to extreme events and improve forecast accuracy. Additionally, we examine the turbulent dynamics generated by solitons in these systems, providing explanations for their onset and evolution. Our analysis offers new insights into chaotic behavior. Finally, we propose applying these methods to real-world wind dynamics to enhance forecasting and deepen understanding of chaotic natural systems.

Keywords: complex systems, spatiotemporal chaos, turbulence, machine learning, artificial neural networks, information theory

---

Optimisation de l'Horizon de prédictibilité des Evènements Extrêmes par «Deep Learning»

Résumé

Des vagues scélérates aux vents violents, les événements extrêmes peuvent perturber les systèmes naturels et les activités humaines sans avertissement. Bien que ces événements semblent imprévisibles, ils émergent souvent des dynamiques complexes des systèmes chaotiques, en particulier du chaos spatiotemporel, où des motifs se déploient dans le temps et l'espace. Dans cette thèse, nous étudions les événements extrêmes dans des systèmes optiques, en nous concentrant sur un résonateur à fibre optique modélisé par l'équation de Lugiato-Lefever. Ce système offre un environnement contrôlé pour analyser les comportements chaotiques à l'origine de ces phénomènes. Les récents progrès en apprentissage automatique, notamment avec les réseaux de neurones, offrent de nouveaux outils pour prédire les dynamiques chaotiques. Cependant, la prévision à long terme reste difficile en raison de l'imprévisibilité inhérente au chaos. Nous proposons d'étendre l'horizon de prédiction en utilisant des méthodes de théorie de l'information, telles que l'entropie de transfert, pour identifier les régions locales contribuant aux événements extrêmes et améliorer la précision des prévisions. En outre, nous examinons les dynamiques turbulentes générées par les solitons dans ces systèmes, en proposant des explications sur leur apparition et leur évolution. Notre analyse offre de nouvelles perspectives sur le comportement chaotique. Enfin, nous proposons d'appliquer ces méthodes aux dynamiques du vent en situation réelle pour améliorer les prévisions et approfondir la compréhension des systèmes chaotiques naturels.

Mots clés : systèmes complex, chaos spatiotemporel, turbulence, apprentissage automatique, réseaux de neurones artificiels, théorie de l'information

---

Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM)
CNRS UMR 8523 – Université de Lille – Bâtiment P5 campus Cité Scientifique – 2 Avenue Jean Perrin – 59655 Villeneuve d'Ascq Cedex – France