# Expanding CCM Applicability: Overcoming Challenges in Causal Discovery

*A REPORT*

*submitted by*

**KALASH VERMA**

**(ME18B052)**

*for the final review*

*of*

**BACHELOR OF TECHNOLOGY**

MECHANICAL ENGINEERING

and

**MASTER OF TECHNOLOGY**

COMPLEX SYSTEMS AND DYNAMICS

*Under the guidance of*

**DR. ARUN K. TANGIRALA**

Professor
Department of Chemical Engineering

**INDIAN INSTITUTE OF TECHNOLOGY MADRAS**

**June 2023**

# THESIS CERTIFICATE

This is to certify that the thesis titled **Expanding CCM Applicability: Overcoming Challenges in Causal Discovery**, submitted by **Kalash Verma (ME18B052)**, to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelor of Technology** and **Master of Technology** is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

**Prof. Arun K. Tangirala**
Research Guide
Professor
Dept. of Chemical Engineering
IIT Madras, 600 036

Place: Chennai
Date: June 5, 2023

# ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Professor Arun K. Tangirala for his invaluable guidance, support, and constant encouragement throughout the course of my master's project. His expertise, insightful feedback, and dedication have been instrumental in shaping my research and academic journey.

I would also like to extend my heartfelt appreciation to all my professors at IIT Madras, who have played a significant role in shaping my knowledge and intellectual growth over the past five years.

Furthermore, I would like to acknowledge the support and encouragement I have received from my friends. Their camaraderie, stimulating discussions and willingness to lend a helping hand has made this journey all the more enjoyable and fulfilling.

Lastly, I am deeply grateful to my parents for their unconditional love, constant encouragement, and unwavering belief in my abilities. Their sacrifices and unwavering support have been the cornerstone of my success.

# ABSTRACT

KEYWORDS: Causality Analysis, Convergent Cross Mapping, Non-Linear systems, Synchronization Likelihood, Generalized Synchrony, Multivariate Convergent Cross Mapping, Coupling Strength, Complex Systems

In various domains such as medicine, ecology, climatology, economics, business, and engineering, the ability to infer causal structures directly from time series measurements holds significant potential. Convergent Cross Mapping (CCM) emerges as a promising tool for inferring the causal structure of complex non-linear dynamical systems. CCM builds upon the principles of Takens' embedding theorem and its extensions. Despite recent advancements, CCM still faces several limitations that restrict its application. This research aims to investigate and address some of these challenges. A primary limitation identified in the existing literature is the use of CCM in the presence of strong unidirectional coupling. However, there is currently no established method to quantify the magnitude of strong unidirectional coupling and determine whether CCM can be reliably applied to a given dataset. In this study, we propose Synchronization Likelihood as a metric to assess the reliability of the CCM dataset. Furthermore, we propose an alternative hypothesis that suggests the applicability of CCM can be better evaluated by considering the level of synchrony rather than the strength of coupling alone. Additionally, we explore the limitations and applicability of Multivariate CCM, considering its potential in differentiating between direct and indirect causal links. Furthermore, we investigate the effect of noise on the CCM framework. By addressing these limitations and proposing novel approaches, this research contributes to enhancing the understanding and utilization of CCM for inferring causal structures from time series data.

# TABLE OF CONTENTS

# LIST OF FIGURES

# ABBREVIATIONS

| | |
|---|---|
| **IITM** | Indian Institute of Technology, Madras |
| **CCM** | Convergent Cross Mapping |
| **CMS** | Cross Map Skill |
| **SSR** | State Space Reconstruction |
| **Sl** | Synchronization Likelihood |
| **GS** | General Synchrony |
| **SNR** | Signal to Noise Ratio |
| **PAI** | Pairwise Asymmetric Inference |
| **C** | Coupling Strength |

# CHAPTER 1

# INTRODUCTION

Causality refers to the relationship or connection between a cause and an effect. It is a study of how one thing (a cause) influences or leads to the production of another thing(an effect). Understanding cause and effect is crucial for modelling, comprehending, and modifying systems to our advantage. By identifying the causal relationships, we can determine which lever to manipulate in order to bring about a specific change in the system or predict the consequences of altering underlying assumptions. The study of causality has found practical applications in various domains such as medicine, ecology, climatology, economics, business, and engineering, among others.

Traditionally, researchers identified causal relationships through controlled experiments. However, practical or ethical constraints often render experimentation infeasible. Consequently, it becomes imperative to develop frameworks that can unveil causality from observational data. This process, known as causal discovery, involves inferring causal structures from observed information. With the advent of abundant data, causal discovery has become a highly relevant and appealing field.

Many researchers have attempted to develop causal discovery frameworks achieving varying degrees of success. Nevertheless, there is no universally applicable framework that can be used for causal discovery. Causal discovery methods can be further categorised into two classes model-based and model-independent methods. Currently, researchers employ diverse approaches tailored to their specific requirements, data types, and applications. This work focuses on one such framework, namely Convergent Cross Mapping (CCM), introduced by Sugihara *et al.* (2012). We investigate several limitations of the CCM methodology that impede its widespread adoption among practitioners and propose potential solutions. In particular, we address the challenges associated with strongly coupled systems and develop a reliable metric for assessing the applicability of CCM. Additionally, we explore the impact of noise on Multivariate CCM.

The structure of the thesis is as follows: Section 2 provides a basic overview of theoretical building blocks. In Section 3, we delve into the literature, offering a concise account of the development of CCM, highlighting its limitations, and identifying areas that require further research. The methodology used in this work is outlined in Section 4. In Section 5, we present our findings and results. In section 6 we analyze the results and their implications. Finally, Section 7 concludes the thesis by discussing the derived conclusions from this study and suggesting future research directions.

# CHAPTER 2

# THEORETICAL FOUNDATIONS

## 2.1 Non-Linear Dynamical Systems

CCM (Convergent Cross Mapping) is a framework developed specifically for analyzing causality in nonlinear deterministic dynamical systems. In our study, we will focus exclusively on these types of systems. Dynamical systems are characterized by their evolution over time, which is governed by a set of equations. These equations can be expressed either as differential equations or iterated maps. Non-linearity in the governing equations implies the presence of higher-order terms. Deterministic behaviour means that the system's evolution solely depends on its parameters and initial values. These systems can be described by the following equations:

$$\begin{aligned} \frac{d}{dt}\mathbf{x}(t) &= \mathbf{f}(\mathbf{x}(\mathbf{t}), P) \\ \mathbf{x}_{k+1} &= \mathbf{F}\left(\mathbf{x}_k, P\right) \end{aligned} \tag{2.1}$$

where $\mathbf{x}(t)$,$\mathbf{x}_k \in \mathbb{R}^n$ are states of the system and $f, F$ govern the evolution of the states in time, Additionally $P$ are the set of parameters that characterize the system.

Studying nonlinear dynamical systems poses significant challenges, as they cannot be easily decomposed into simpler parts, and superposition principles do not apply. Consequently, we rely on topological and geometric tools to investigate these systems. A crucial concept in the analysis of such systems is the phase space or state space. State space is the space spanned by all possible states of the dynamical system.

For dissipative systems, attractors play a crucial role in understanding the system's behaviour. Attractors are subsets of the phase space where the system tends to evolve, regardless of its initial starting position. The structures of attractors can vary widely, ranging from simple points or curves to complex fractal sets or manifolds. To gain a better understanding of nonlinear dynamical systems and chaos, I recommend referring to Strogatz (2018), which is an excellent introduction on the subject.

## 2.2 State Space Reconstruction

In practical scenarios, we often do not have direct access to all the phase state variables that define the state space of a dynamical system. Instead, we typically work with measurements of some observable property, usually in the form of a time series. These measurements can be viewed as a function that projects the system's states onto lower dimensions. While the measurements are typically one-dimensional, the states themselves exist in a higher-dimensional space. It is crucial to recognize that reconstructing the complete state space solely based on the measurements is not feasible. State space reconstruction (SSR) refers to the technique used to obtain a reconstructed phase space that is in some sense equivalent to the original state space. Various methods have been developed for state space reconstruction, but in this study, our focus will be on the delay coordinate embedding approach.

Takens' theorem (Takens, 1981) provides the foundation for state space reconstruction. This theorem was later extended in Sauer *et al.* (1991), Stark (1999) and Stark *et al.* (2003). In simple terms, Takens' theorem states that given a time series measurement X, the delay coordinate map represented by $M' = [X(t), X(t-\tau), X(t-2\tau)..., X(t-(2m-1)\tau)]$ can serve as an embedding of the original state space of the dynamical system, provided that the embedding dimension m is greater than $2d$ where d is the dimension of the underlying manifold. The resulting reconstructed state space is often referred to as a shadow manifold. It is important to note that not all measurements will lead to successful reconstruction of the state space, and there are certain conditions imposed on the measurement functions (for more details, refer to Takens' original paper). Theoretically, the delay coordinate map will yield a valid embedding as long as $m > 2d$. However, in practice, we typically do not know the exact dimension of the system, and selecting an appropriate value for m becomes a crucial consideration. Similarly, while the embedding theorem does not impose any restrictions on the choice of time delay parameter $\tau$, practical considerations require us to select an appropriate $\tau$ value to ensure reliable embedding.

In certain cases, we may have access to multiple sets of measurements. In such situations, relying solely on a single time series observation for state space reconstruction does not allow us to fully exploit all available information. To address this, Deyle and Sugihara (2011) extended the embedding theorem to enable embeddings using multiple time series. This delay coordinate map can be mathematically represented as $M' = [x_1(t - \tau_{11}), x_1(t - \tau_{12}, ....x_1(t - \tau_{1m_1}, x_3(t - \tau_{21}, ...., x_p(t - \tau_{pm_p}]$ where $x_1, x_2....x_p$ are p scalar time series measurements, $\tau_{ij}$ corresponds to the $j^{th}$ lag for $x_i$ and $m_1, m_2....m_p$ are number of lagged coordinates of $x_1, x_2....x_p$ respectively.

This extension allows us to incorporate multiple time series measurements into the state space reconstruction process, enhancing our ability to capture the underlying dynamics of the system. By including information from multiple measurements and their respective time lags, we can obtain a more comprehensive representation of the system's behaviour in the reconstructed phase space. This ability is especially useful when we have multiple observations of a system, but each time series is limited in its length.

## 2.3 Synchrony

Dealing with synchrony is the central focus of this work, and therefore, we will explore the concept of synchronization in more detail. Specifically, we will delve into the understanding of generalized synchronization (GS) as it plays a crucial role in the applicability of CCM. Here, it is also important to discuss the relationship and difference between coupling strength.

Synchrony loosely refers to the phenomenon of multiple elements or components in a system exhibiting coordinated behaviour or activity. Synchrony can be manifested in many forms, such as synchronized oscillations or rhythmic patterns.

In case of GS, the response signal is uniquely determined by the driving variable. Mathematically, generalized synchronization implies that there exists a functional dependence of the response variable (denoted by $y$) on the driving variable (denoted by $x$) that is $y(t) = \phi(x(t)$ (Rulkov *et al.*, 1995).

Coupling strength, on the other hand, pertains to the strength or intensity of the interactions or connections between different variables. It quantifies the influence that one variable has on another, determining how strongly they affect each other's behaviour or dynamics. There is no concrete mathematical definition of coupling strength.

Generally, increasing coupling strength between systems can influence the emergence or maintenance of synchrony. When the coupling is weak, the two systems can behave more independently, leading to a lack of synchrony. As coupling strength increases, the effect of the driving system on the response variable will increase, thus increasing the likelihood of synchronization. However, this may not always be the case, and in some cases, high coupling strength may result in disruption of synchronization.

### 2.4 Convergent Cross Mapping

In CCM, we try to recover and estimate the value of one variable from the historical data of another variable. This process, known as cross-mapping, involves using the historical data of one variable, say $Y$ to reconstruct and estimate another variable say $X$. By measuring the reliability of this cross-mapping process, CCM tests for causation between the variables.

For a dynamical system, two-time series variables are said to be causally linked if they share a common attractor manifold; that is they belong to the same dynamical system. This means each of the variables contains information about the other variable, courtesy of embedding theorems. These variables show bidirectional causality.

In the case of unidirectional causality, where one variable (e.g., $X$) is driving or causing another variable (e.g., $Y$), the information of the driving variable $X$ is encoded in the effect variable $Y$. This allows us to recover the driving variable $X$ from the effect variable $Y$ using CCM. However, CCM has a limitation in distinguishing between bidirectional causality and strong unidirectional causality that leads to synchrony (Ye *et al.*, 2015). When there is strong unidirectional coupling between variables, the dynamics of the effect variable Y become dominated by the driving variable X. As a result, the system collapses to the dynamics of the driving variable alone. In this scenario, CCM may falsely infer bidirectional causality between the variables, even though the causal link is unidirectional. To address this limitation, Ye *et al.* (2015) proposed considering different lags in the cross-mapping process, allowing the distinction between bidirectional and strong unidirectional causality. Instead of directly cross-mapping $Y(t)$ to $X(t)$, they suggested cross-mapping $Y(t)$ to $X(t+l)$ for reasonable lag values $l$. The hypothesis is that if $X$ forces $Y$ with some time delay, or even instantaneously, the current state of $Y(t)$ will better predict past values of $X$. This would mean maxima of cross-map skill will happen for some $l < 0$.

The process of CCM analysis can be divided into three main steps: data preprocessing, pairwise CCM analysis, and multivariate CCM analysis. Data preprocessing involves testing the suitability of the data for CCM and applying necessary transformations, such as denoising and detrending, to make it suitable for analysis.

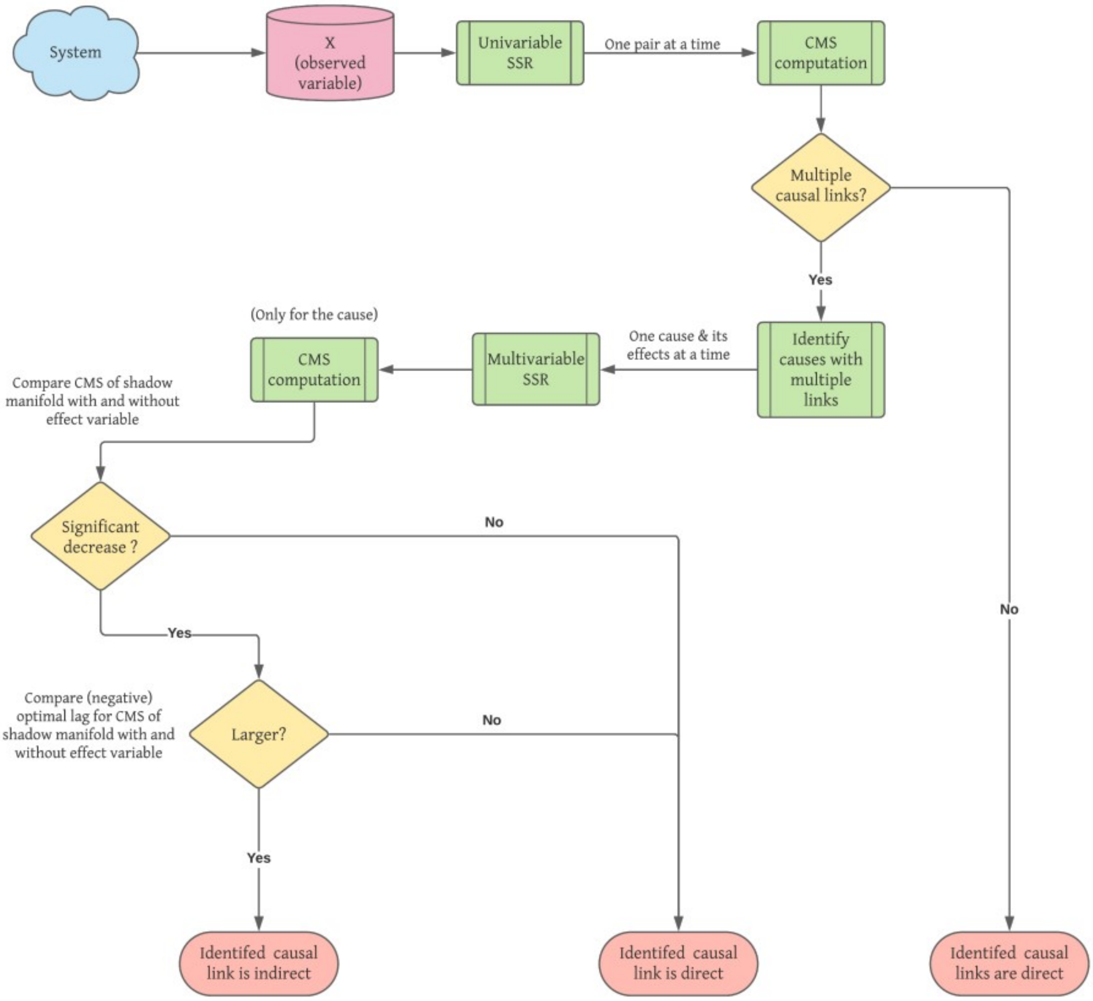

System
X (observed variable)
Univariable SSR
One pair at a time
CMS computation
Multiple causal links?
Yes
No
(Only for the cause)
CMS computation
Multivariable SSR
One cause & its effects at a time
Identify causes with multiple links
Compare CMS of shadow manifold with and without effect variable
Significant decrease ?
No
Yes
Compare (negative) optimal lag for CMS of shadow manifold with and without effect variable
Larger?
No
Yes
No
Identifed causal link is indirect
Identifed causal link is direct
Identifed causal links are direct

Figure 2.1: Flowchart summarizing steps for CCM analysis(Nithya and Tangirala, 2021)

### 2.4.1 Pairwise CCM Analysis

In pairwise CCM analysis, all pairs of time series measurements are taken, and CCM analysis is performed to determine the causal links between these variables. It is important to note that the pairwise CCM analysis cannot distinguish between direct and indirect links, and the causal structure obtained represents a transitive closure graph of the original causal interaction graph. The difference between the causal interaction graph and its transitive closure graph is illustrated in the figure 2.2.

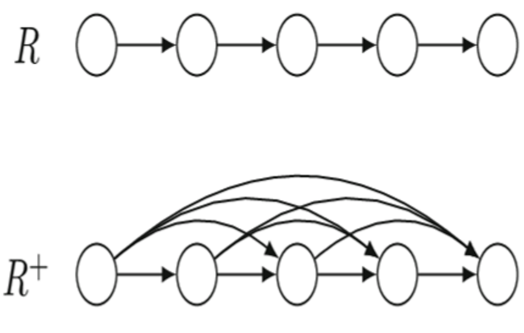


Figure 2.2: $R^+$ is the transitive closure graph of $R$

### 2.4.2 Steps for pairwise CCM analysis

1. *Choose a pair of time series measurements ($X$ and $Y$) between which the nature of causal link is to be identified.* Between a pair of variable, $X$ and $Y$ four types of causal links can exist, bidirectional link ($X \leftrightarrow Y$), unidirectional causal link with $X$ driving $Y$, ($X \rightarrow Y$) or $Y$ driving $X$ ($Y \rightarrow X$ ) or no causal link ($X$ $Y$).

2. *Reconstruction of shadow manifold* $\mathbf{M_x}$ *and* $\mathbf{M_x}$*. Here we also choose the library size L <= N (data length).* The shadow manifold is given by $\mathbf{M_x}$ is the set of lagged coordinate vectors given by

$$\bar{x}(t) = [X(t), X(t-\tau), X(t-2\tau)..., X(t-(2m-1)\tau)]$$

where,$X(t)$ denotes time series measurement at time $t$, $m$ is embedding dimension and $\tau$ is delay time . $t$ ranges from $t = 1 + (m-1)\tau$ to $t = L$ where $L$ is the library length. Similarly, $\mathbf{M_y}$ can also be reconstructed.

3. *Generate a Cross Mapped Estimate of $Y$, denoted by $Y'|\mathbf{M_x}$,and similarly generate cross mapped estimate of X($X'|\mathbf{M_y}$).* To generate a cross-mapped estimate $Y'(t)|\mathbf{M_x}$, first we locate contemporaneous lagged coordinate vector on $\mathbf{M_x}$ represented by $\bar{x}(t)$. Next we find $m+1$ neighbours of $\bar{x}(t)$ and denote them by $\bar{x}(t_i)$, ordered from nearest to farthest.$(t_i)$ is the time index of the $i^{th}$ nearest neighbour. We can estimate $Y'(t)|\mathbf{M_x}$ using the equation

$$Y'(t) \mid \mathbf{M_X} = \sum w_i Y(t_i) \quad i = 1 \ldots m+1 \tag{2.2}$$

where weights $w_i$ is defined as

$$w_i = u_i / \sum u_j \quad j = 1 \ldots m+1 \tag{2.3}$$

where

$$u_i = exp{-}d[\bar{x}(t), \bar{x}(t_i)]/d[\bar{x}(t), \bar{x}(t_1)] \tag{2.4}$$

and $d[\bar{x}(t), \bar{x}(t_i)]$ is the L2 norm between two vectors. We can find $Y'|\mathrm{M_x}$ by finding cross map estimate for all $t = 1 + (m-1)\tau \ldots L$. $X'|\mathrm{M_y}$ can be found analogously.

4. *Computing the Cross Map Skill for various Library lengths.* Cross Map Skill can be computed either by using correlation or coefficient of determination between $Y$ & $Y'|\mathrm{M_x}$, and $X$ and $X'|\mathrm{M_y}$ for different library sizes L. To infer a causal link from $X$ to $Y$ ($X \to Y$), the CMS should be significant, increase with the increase in L and show convergence.

5. *Performing extended CCM analysis if a bidirectional link is observed.* To perform extended CCM analysis we cross map $Y(t)$ to $X(t+l)$ for different lags $l$ and similarly cross map $X(t)$ to $Y(t+l)$, generally for maximum library size L. To infer a causal link from $X$ to $Y$, the optimal CMS for cross-mapping of $Y(t)$ to $X(t+l)$ should be at $l < 0$.(Ye *et al.*, 2015)

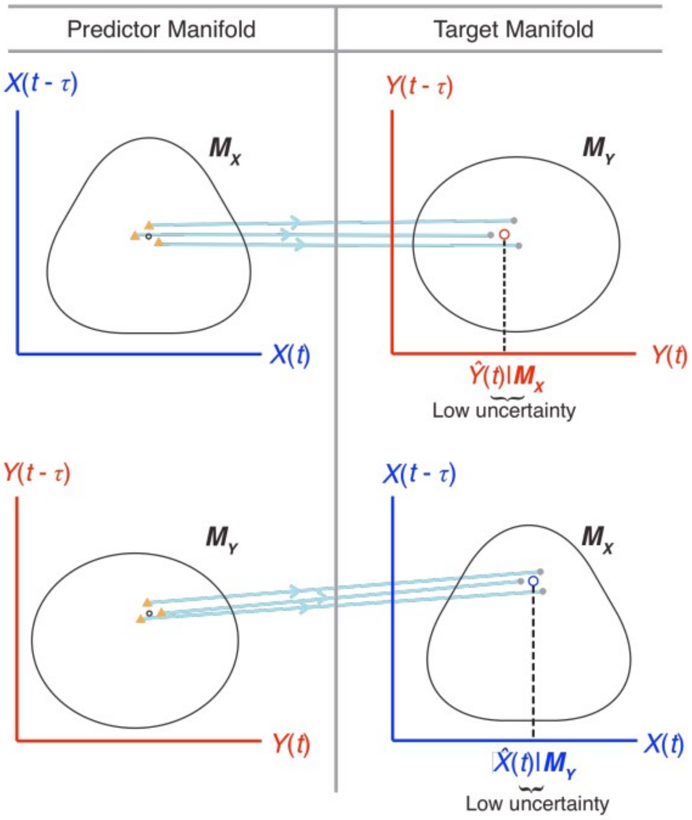


Figure 2.3: Visualization of Cross mapping in the presence of bidirectional causality.

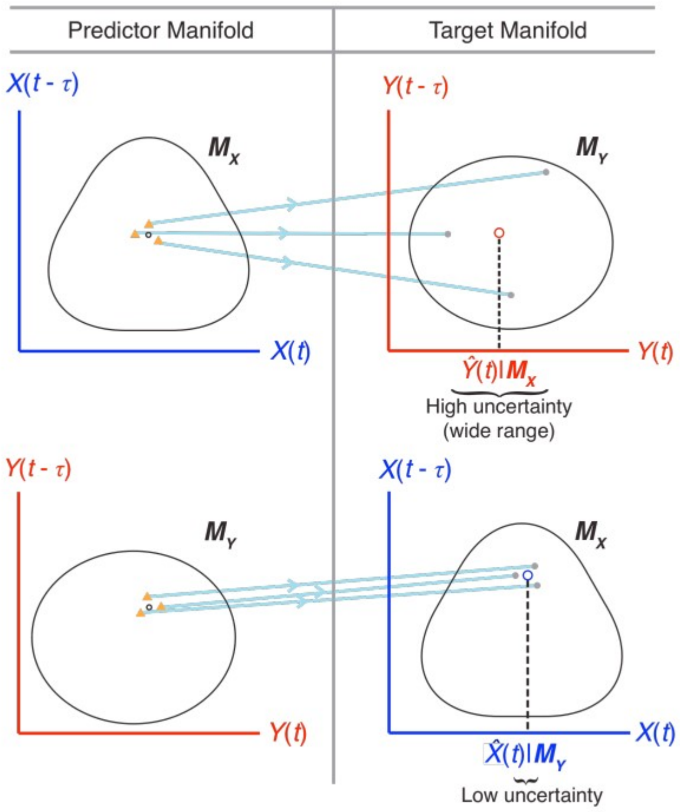


Figure 2.4: Visualization of Cross mapping in the presence of unidirectional causality.

### 2.4.3 Multivariable CCM Analysis

Multivariable CCM analysis is done to distinguish between direct and indirect links. We first identify variables which act as causes of multiple effects. We analyse each set of cause variables along with its set of effect variables using multivariable SSR method. For example, let X act as a cause to Y and Z variables.We first construct $\mathbf{M_y}$,$\mathbf{M_z}$,$\mathbf{M_{yz}}$ shadow manifold. Then we calculate CMS for X using the three-manifolds. We compare the CMS of the shadow manifold with and without the effect variable. If there is a significant decrease, it indicates an indirect link between the cause and removed effect variable. It is also necessary to check whether the optimal lag for the indirect effect is more negative than for the direct one.(Nithya and Tangirala, 2021)

## 2.5 Synchronization Likelihood (SL)

The Synchronization Likelihood (SL) metric, introduced by Stam and van Dijk (2002), serves as a measure to quantitatively assess the level of general synchrony between coupled dynamical systems. This metric adopts a state space-based approach, leveraging the concept of embedding vectors to evaluate synchronization likelihood.

The SL method involves calculating the likelihood of simultaneous auto-recurrence of embedding vectors within the phase space of the dynamical systems. This calculation can be summarized in the following steps:(Stam and van Dijk, 2002; Khanmohammadi, 2017; Montez *et al.*, 2006).

1. The first step in the Synchronization Likelihood (SL) method is to construct embedding vectors, which are similar to those described in the State Space Reconstruction section. For a time series data represented by $x_i^k$ where $k = 1, 2...K$, $K$ represents the number of different time series measurements and $i = 1, 2, ...N$, $N$ denoting length of time series. The embedding vector is

$$X_i^k = x_i^k, x_{i+l}^k \dots, x_{i+(m-2)l}^k, x_{i+(m-1)l}^k$$

   with l and m as delay time and embedding dimension respectively.

2. In the second step of the Synchronization Likelihood (SL) method, we calculate the likelihood of simultaneous auto-recurrence of embedding vectors, for two-time series data points at index $i$. This likelihood is computed using the following equation2.5.

$$SL_i^{k1,k2} = \frac{\sum_{j\in[\pm w2]\notin[\pm w1]} \Theta\left(\varepsilon_i^{k_1} - \left\|X_i^{k_1} - X_{i+j}^{k_1}\right\|_2\right) \Theta\left(\varepsilon_i^{k_2} - \left\|X_i^{k_2} - X_{i+j}^{k_2}\right\|_2\right)}{n_{\text{rec}}} \tag{2.5}$$

   $\varepsilon$ is called similarity threshold and is calculated by setting $P$ in equation 2.6 to some $P_{ref} << 1$.

$$P_{X_i^k}^{\varepsilon_i^k} = \frac{1}{2\left(w_2 - w_1\right)} \sum_{j\in[\pm w2]\notin[\pm w1]} \Theta\left(\varepsilon_i^k - \left\|X_i^k - X_{i+j}^k\right\|_2\right) \tag{2.6}$$

   Here $\theta$ is the Heaviside function which takes value for positive values else is 0. $w_1$ and $w_2$ are inclusion and exclusion windows.

$$n_{rec} = P_{ref} * (w_2 - w_1 - 1)$$

3. The overall Sl value is calculated by taking the average over all values of i (time).

# CHAPTER 3

# LITERATURE REVIEW

## 3.1 Development of CCM Framework

Granger causality, proposed by Granger in 1969 (Granger, 1969), is one of the most popular approaches utilized to identify causality between time series. In simple terms, a variable $X$ is considered to granger cause another variable $Y$ if its past values contribute to improving the prediction of $Y$. While Granger causality is effective for stochastic and linear systems, it falls short when applied to deterministic dynamical systems. In such systems, if $X$ causes $Y$, the information from $X$ will already be present in $Y$ as a consequence of Takens' theorem and its extensions(Stark, 1999; Stark *et al.*, 2003). Consequently, if we were to employ the Granger causality definition in these cases, we would not observe an improvement in prediction performance since the causal variable's information is already incorporated in the effect variable. To address this limitation, Sugihara *et al.* (2012) proposed the CCM framework for systems where Granger causality is ineffective. In the paper, the authors noted that the CCM framework is not suitable for systems with strong coupling. In 2015, Ye *et al.* (2015) expanded upon the CCM framework by explicitly considering time lags, enabling the differentiation between strong unidirectional forcing and true bidirectional causality. However, it is crucial to note that the CCM framework, as envisioned by Ye *et al.* (2015), is unable to distinguish between direct and indirect causal links. To overcome this challenge, Nithya and Tangirala (2021) recently proposed a Multivariable CCM approach. In this study, we adopt the multivariable CCM Framework proposed in Nithya and Tangirala (2021). This framework is explained in detail in Chapter 2.

## 3.2 CCM and its Limitations

CCM in the short time after its conception has been scrutinized and several different augmentation of the framework has been proposed by various researchers. There are four major failure modes of CCM that have been discussed in the literature (Bartsev *et al.*, 2021; Yuan and Shou, 2022). We provide a brief summary of the failure modes below:

- **Non-Reverting Continuous dynamics:** Non-reverting dynamics can be thought of as a nonstationarity for State space reconstruction techniques.It characterizes the situation where one time series X maintains a consistent pattern over time (non-reverting), while the other time series Y exhibits continuity with respect to time. In the presence of non-reverting dynamics, the CCM method falsely infers bidirectional causality even when it does not exist.
- **Synchrony:** Synchrony occurs when strong unidirectional forcing is present. In such cases, the dynamics of the forced variable are dominated by the forcing variable, leading the CCM framework to occasionally misjudge bidirectional causality in the presence of a unidirectional link. We will delve deeper into synchrony in this work.
- **Oscillations with Integer Multiple Periods:** In cases where the two time series under analysis have periods which are integral multiples of each other, CCM can incorrectly infer causality when there is none.
- **Pathological Symmetry:** In systems with symmetry, certain measurements may fail to fully reconstruct the state space, resulting in erroneous inferences. This is due to limitations imposed by Takens' Theorem.

Apart from the aforementioned four failure modes, McCracken and Weigel (2014) argues in their paper that the CCM framework does not align with intuitive concepts of driving and is not a reliable indicator of causality. They propose a modified analysis called pairwise asymmetric inference (PAI) as a solution to this issue. However, PAI methods entail certain theoretical challenges that require more rigorous treatment and are beyond the scope of this work. Other modifications, such as using Gaussian Process to enhance attractor reconstruction (Feng *et al.*, 2020*b*) or employing controlled noise injection (Mønster *et al.*, 2017) to improve the reliability of CCM in strongly coupled systems, have been proposed by researchers.

Despite these limitations, Convergent Cross Mapping has been applied to causal discovery in numerous real-world datasets. For instance, Díaz *et al.* (2022) employ the CCM framework to infer spatial patterns of causal relations among key variables in carbon and water cycles, such as gross primary productivity, latent heat energy flux for evaporation, surface air temperature, precipitation, soil moisture, and radiation. Feng *et al.* (2020*a*) apply CCM to cardiotocography signals to discover causal relationships between different cardiotocography features and fetal acidosis. CCM methods is also very popular while studying ecosystems(Yuan and Shou, 2022).

## 3.3 Other Causal Discovery Methods

Several other causal discovery methods have been proposed by researchers to address non-linear systems. Demiralp and Hoover (2003) and Malinsky and Spirtes (2018) have studied causal inference using vector regression, providing alternative approaches in this domain. Additionally, graph model-based methods for causal discovery have been explored and discussed in Glymour *et al.* (2019). Information-theoretic measures have also been utilized for causal discovery, as demonstrated by Keskin and Aste (2020). Furthermore, Runge *et al.* (2019) introduced a method that combines linear or nonlinear conditional independence tests to infer the causality structure from large-scale time series datasets, such as climate systems.

State space reconstruction has also been employed in causal discovery, with various methods discussed in the works of Cummins *et al.* (2015),Ma *et al.* (2014), and Harnack *et al.* (2017). These approaches provide alternative perspectives on leveraging state space reconstruction for causal inference.

## 3.4 Gaps in Literature

The development of the CCM framework is relatively recent, as it was proposed in 2012, spanning just over a decade. Despite its emergence, there are still several gaps and areas that necessitate further research. In this section, we will outline the fundamental steps of the CCM framework and offer a concise overview of the existing gaps in the literature. Figure 3.1 provides a visual summary of the discussion.

### 3.4.1 Data Preprocessing

In CCM analysis, the initial step involves working with a set of time series observations. The first task is to determine if the data is suitable for the CCM framework, or if it requires any transformations to make it compatible. However, our literature review did not yield a comprehensive guide or research specifically focused on data preprocessing methods for CCM. Nonetheless, some researchers have employed various basic techniques to identify unsuitable data (Yuan and Shou, 2022). Data preprocessing typically includes tests aimed at reducing and eliminating noise from the observed data.

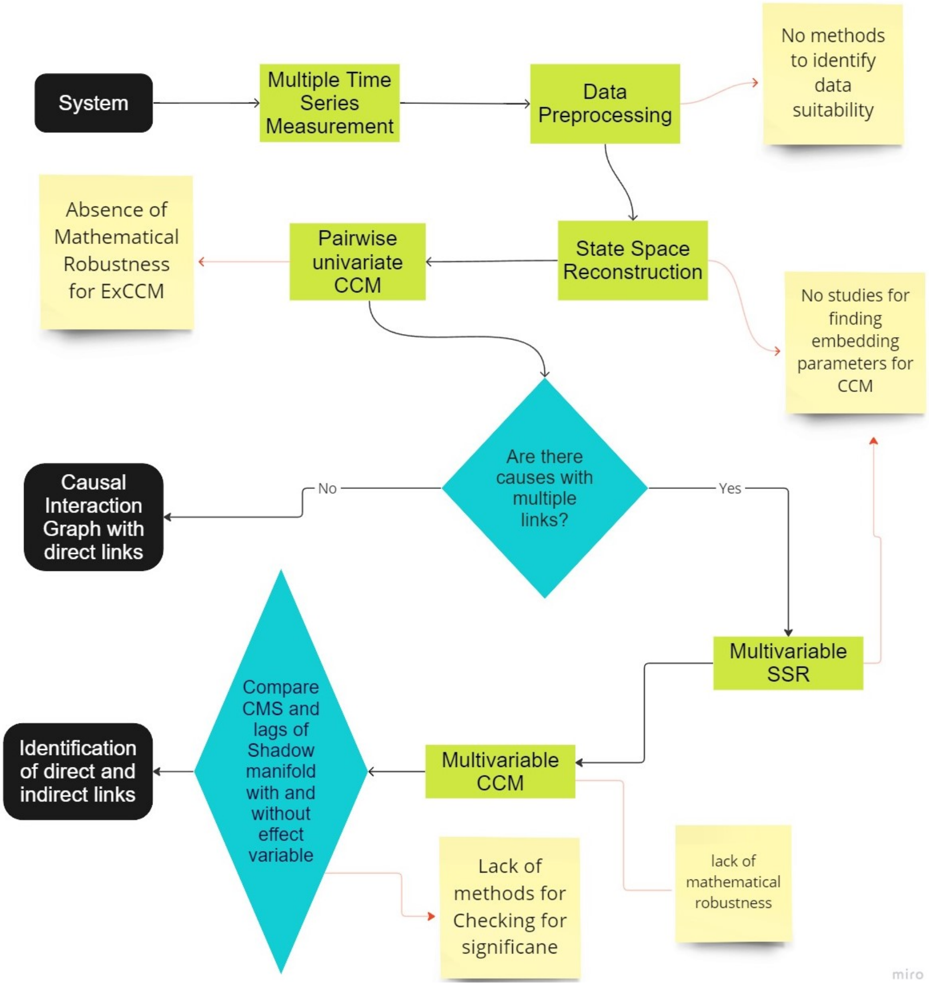

System
Multiple Time Series Measurement
Data Preprocessing
No methods to identify data suitability
Absence of Mathematical Robustness for ExCCM
Pairwise univariate CCM
State Space Reconstruction
No studies for finding embedding parameters for CCM
Are there causes with multiple links?
No
Yes
Causal Interaction Graph with direct links
Multivariable SSR
Compare CMS and lags of Shadow manifold with and without effect variable
Identification of direct and indirect links
Multivariable CCM
Lack of methods for Checking for significane
lack of mathematical robustness
miro

Figure 3.1: Summary of Gaps in literature

Certain studies have explored how noise affects Convergent Cross Mapping (Mønster *et al.*, 2017; Yuan and Shou, 2022). However, we could not locate any literature providing guidance on how to handle noise or a metric to measure noise that could inform practitioners about the applicability of CCM. Similarly, it is also essential to determine the required length of the time series for CCM analysis.

Another challenge pertains to strong coupling, which is not clearly defined and quantified in the literature. This lack of clear definition poses a problem, as practitioners cannot ascertain whether CCM can be applied to a given set of time series data. Addressing this problem is one of the objectives of our research.

### 3.4.2 Univariable State Space Reconstruction

Following data preprocessing, the next step in the CCM analysis is univariable state space reconstruction (SSR). A critical aspect of SSR is the selection of embedding parameters. Incorrect choices of lag and embedding dimension can lead to significant errors in inferences. The determination of these parameters heavily relies on the type, quality, and quantity of data, as well as the intended application. While there is ample literature on finding optimal embedding parameters for prediction purposes(Krakovská *et al.* (2022); hui Lang *et al.* (2021) provide a good review of methods for choosing embedding parameters), further research is needed to understand how the choice of method for finding these parameters affects the causal analysis.

### 3.4.3 Pairwise CCM Analysis

The CCM framework proposed by Sugihara *et al.* (2012) is based on Takens' theorem and its extensions. However, several proposed augmentations, including extension proposed by Ye *et al.* (2015), lack robust mathematical theories to support their claims. Although the method by Ye *et al.* (2015) utilize information-based arguments, there is a dearth of literature explaining why these methods would be effective based on the topological theory underlying the original CCM framework.

### 3.4.4 Multivariate SSR

Unlike univariate SSR, the multivariate state space reconstruction method is relatively new, with its mathematical justification provided in 2011 (Deyle and Sugihara, 2011). Consequently, multivariate SSR remains underexplored compared to univariate SSR. There is significant research potential in exploring, comparing, and developing different schemes of multivariate SSR for various applications, including causal discovery.

### 3.4.5 Multivariate CCM

Multivariate CCM was very recently proposed by Nithya and Tangirala (2021), and thus, the limits of its applicability have not yet been thoroughly tested. Additionally, there is a lack of robust mathematical foundations supporting the multivariate CCM framework.

# CHAPTER 4

# METHODOLOGY

In this chapter, we outline the methods used to answer the research objectives.

## 4.1 Systems Description

For our analysis and to demonstrate our results, we have utilized three distinct dynamical systems: the 2D coupled Henon Map, the 2D coupled Logistic Map, and the 3D coupled Logistic Map. These systems, although simple in nature, provide us with the flexibility to manipulate their parameters and investigate various phenomena.

### 4.1.1 2D Coupled Henon Map

The 2D coupled Henon Map is a discrete-time system that exhibits chaotic behaviour. It consists of two variables, denoted as x and y, which evolve iteratively according to the equations given in 4.1.

$$\begin{aligned} X(t+1) &= a_1 - X(t)^2 + b_1 X(t-1) \\ Y(t+1) &= a_1 - (CX(t) + (1-C)Y(t))Y(t) + b_2 Y(t-1) \end{aligned} \tag{4.1}$$

The coupled Henon map is coupled through the variable C and has a unidirectional causal link $X \to Y$ when $C > 0$. We can change the coupling between $X$ and $Y$ by varying the value of C to study how coupling strength affects CCM.

### 4.1.2 2D Coupled Logistic Map

Similar to coupled Henon Map, the 2D coupled Logistic Map is another discrete-time system that demonstrates chaotic behaviour. The system evolves according to the set of equations represented in 4.2.

$$
\begin{aligned}
X(t+1) &= X(t)\,[a_{11} - (a_{11}X(t) + a_{12}Y(t)] \\
Y(t+1) &= Y(t)\,[a_{21} - (a_{21}X(t) + a_{22}Y(t)]
\end{aligned}
\tag{4.2}
$$

Here the parameter $a_{12}$ governs the coupling from $Y$ to $X$ and $a_{21}$ governs the coupling from $X$ to $Y$. If $a_{12} > 0$, Y will causally affect X and similarly, if $a_{21} > 0$, X will causally affect Y.

### 4.1.3 3D coupled Logistic Map

3D coupled Logistic Map extends the previous 2D system. Equations 4.3 are the governing equation.

$$
\begin{aligned}
X(t+1) &= X(t)\,[a_{11} - (a_{11}X(t) + a_{12}Y(t) + a_{13}Z(t)] + \eta \\
Y(t+1) &= Y(t)\,[a_{21} - (a_{21}X(t) + a_{22}Y(t) + a_{23}Z(t)] + \eta \\
Z(t+1) &= Z(t)\,[a_{31} - (a_{31}X(t) + a_{32}Y(t) + a_{33}Z(t)] + \eta
\end{aligned}
\tag{4.3}
$$

We use this system to study Multivariate CCM.

## 4.2 Synchronization Likelihood

To assess the viability of Synchrony Likelihood (SL) as a metric for evaluating the applicability of CCM, we conducted CCM analysis across various coupling strengths. Additionally, SL values were calculated for each of these coupling strengths, allowing us to examine cases where correct and incorrect inferences were made and compare them with the corresponding SL values.

By performing CCM analysis, we aimed to understand the effectiveness of SL in capturing the underlying dynamics of the coupled systems. By comparing the SL values with the outcomes of the CCM analysis, we could determine the range of SL values associated with accurate and inaccurate inferences.

For calculating the SL metric between two-time series data, we utilized the implementation provided by the eeglib Python library, as described in Cabañero-Gomez *et al.* (2021). This library offers a reliable method to calculate the SL metric, which quantifies the level of synchronization between the time series.

The parameters required for SL calculation were determined using the approach outlined in Montez *et al.* (2006).

### 4.3 Noise analysis

In the context of a dynamical system, noise can be categorized into two main groups: measurement noise and process noise. Measurement noise refers to errors or disturbances in the measured data that are independent of the underlying system dynamics. On the other hand, process noise, also known as dynamical noise, arises from small random perturbations introduced into the system at each step.

Consider a dynamical system described by the equation:

$$\mathbf{x_{n+1}} = \mathbf{F}(\mathbf{x_n})$$

In the presence of measurement noise, the observed data can be represented as:

$$s_n = s(\mathbf{x_n}) + \eta_n$$

where $s(\cdot)$ is an observation function that maps the system's states to the real line, and $\eta_n$ represents the measurement noise at time step $n$. The series of $\eta_n$ values constitutes the measurement noise.

On the other hand, dynamical or process noise can be incorporated into the system dynamics as follows:

$$\mathbf{x_{n+1}} = \mathbf{F}(\mathbf{x_n} + \boldsymbol{\eta_n})$$

Here, $\eta_n$ represents the dynamical noise that affects the system's state transition at each time step.

In this work, the focus is on investigating the effect of measurement noise on the ability of CCM analysis to correctly identify whether a causal link between variables is direct or indirect. The measurement noise is modelled as Gaussian white noise. Gaussian white noise is characterized by its randomness, with each sample drawn independently from a Gaussian distribution with zero mean and constant variance. The analysis is done for various SNR(Signal to noise ratio). SNR is defined as ratio of variance of signal to variance of noise.

By studying the impact of measurement noise on CCM analysis, we can gain insights into the robustness and reliability of the method in real-world scenarios where noise is present.

### 4.4 CCM analysis

The CCM framework proposed by Nithya and Tangirala (2021) is used for the analysis. The framework is outlined in Chapter 2.

For univariate state space reconstruction, the self-mutual information method is used to find the delay time $\tau$, and the method of false nearest neighbour is employed to determine the embedding dimension $m$.

Multivariate state space reconstruction is performed using the scheme given in Vlachos and Kugiumtzis (2009).

To conduct a significance test, we employ the surrogate data method, which involves generating surrogate data samples through the process of phase randomization. Subsequently, we determine the 100(1-$\alpha$)% quantile from the surrogate data samples and compare it with the CMS value.

# CHAPTER 5

# RESULTS AND INFERENCES

## 5.1 SL: A metric to assess the applicability of CCM

In our study of the systems, we observed that the Synchrony Likelihood (SL) range of [0,1] can be roughly divided into three distinct regions for unidirectional forced systems.

The first region (Region I) corresponds to SL values below 0.15. Within this region, we found that the CCM analysis fails as the Convergent Cross Mapping (CCM) values are not statistically significant. Consequently, we infer no causality.

The second region (Region II) occurs when SL values range between 0.15 and 0.6. Within this range, the CCM method performs correctly, allowing us to make accurate inferences regarding the existence of causal links between the systems.

However, for SL values exceeding 0.6, we observed that the CCM analysis fails. In this region, there is a tendency to identify unidirectional causal links as bidirectional incorrectly. We will call this region, where CCM fails due to a high amount of synchrony Region III. We illustrate our findings below.

### 5.1.1 Case 1: Coupled Identical Henon System

The system is described using the equation 4.1. The parameters are $a_1 = 1.4, b_1 = b_2 = 0.3$. $C$ is the coupling strength and is varied from 0.01 to 0.8. For $C > 0.75$ the systems are in complete synchrony. The system is called identical because $b_1 = b_2$

In Figure 5.1, we observe a general trend where the SL value tends to increase as the coupling strength increases. However, it is important to note that this relationship is not always the case, and an increase in coupling strength does not necessarily result in an increase in SL.

For very weak coupling strength $c < 0.15$ and low values of $SL < 0.07$ the CMS for both the cross-maps is very low and not statistically significant.

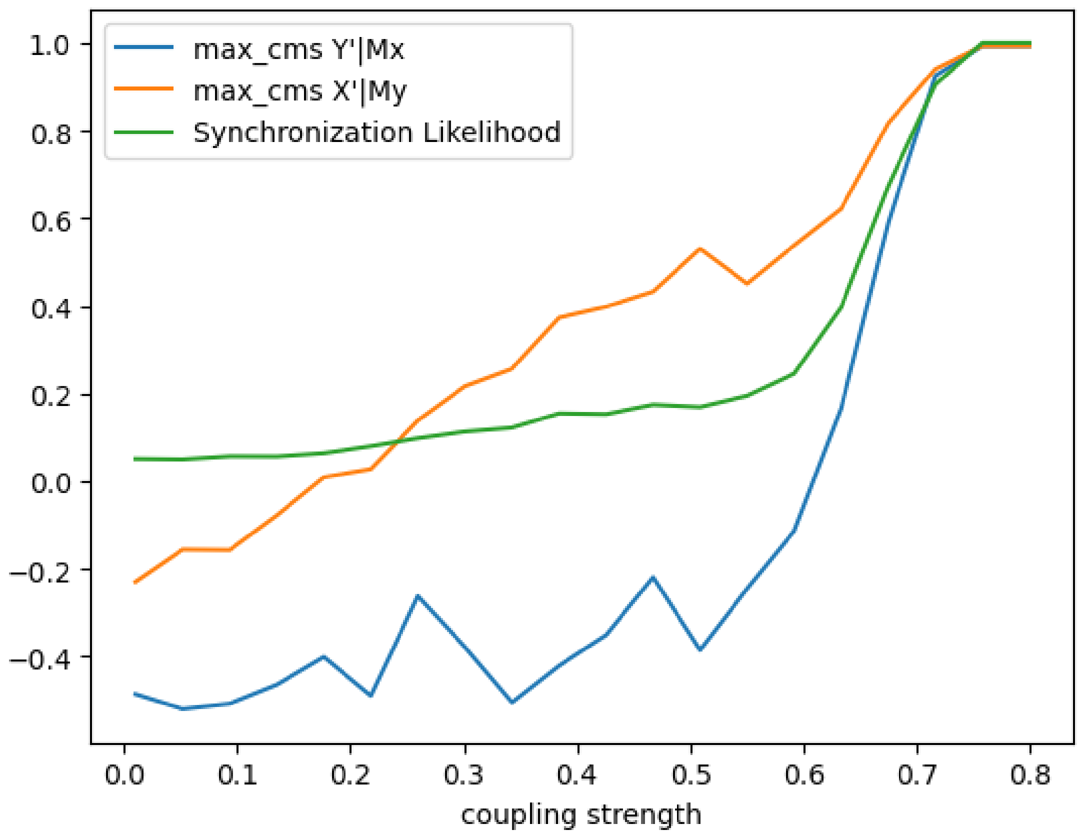


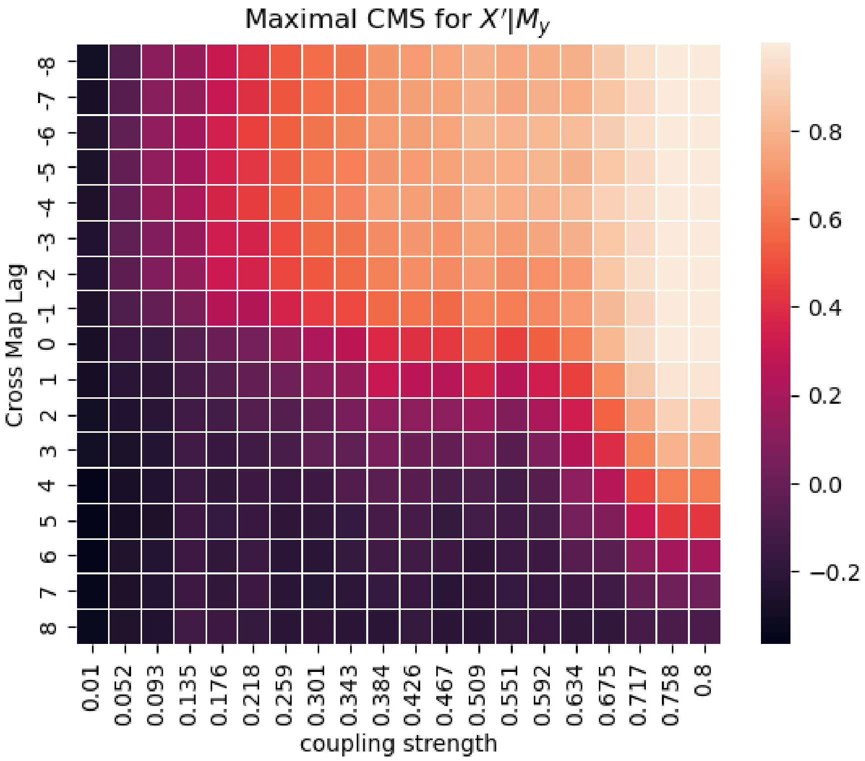
Figure 5.1: Plot of maximum CMS achieved for $X'|M_y andY'|M_x$ with increasing coupling strength for identical henon system). Sl vs coupling strength is also plotted

When the coupling strength is between 0.15 and 0.5, with corresponding SL values ranging from 0.07 to 0.5, we can observe from Figure 5.1 that the maximum CMS becomes significant for $X'|M_y$, allowing us to correctly infer that X causes Y and that there is no causal link from Y to X. This observation is further supported by the heat maps shown in Figure 5.2a, where the optimal CMS for $X'|M_y$ occurs at negative lags.

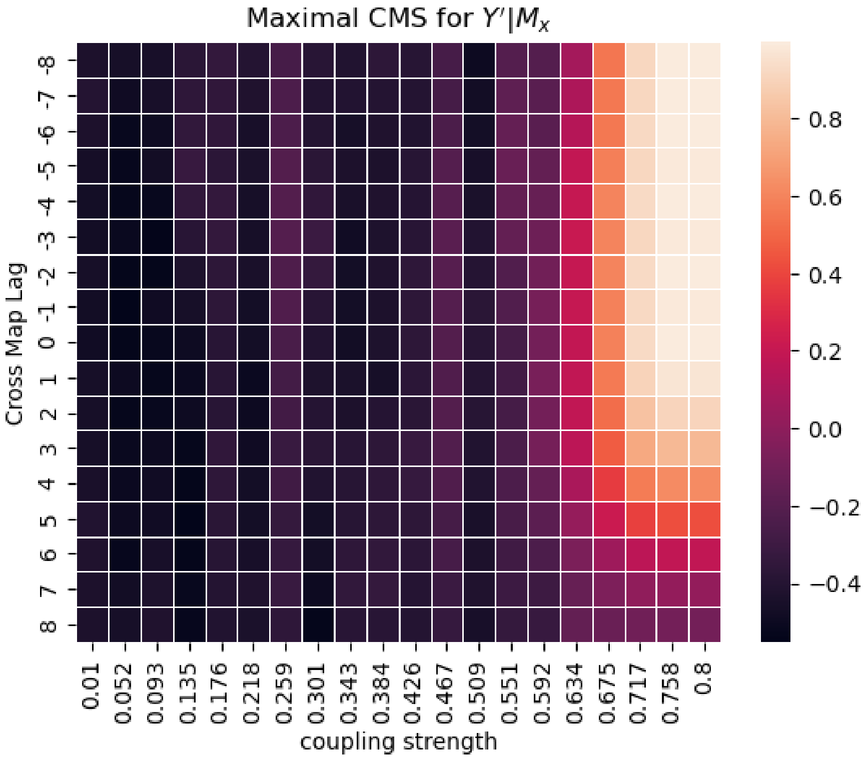
For coupling strengths between 0.5 and 0.67, with SL values ranging from 0.5 to 0.67, the maximum CMS for $Y'|M_x$ starts to become significant, as observed in Figure 5.1. However, it is noteworthy that the optimum lag at which the maximum CMS for $Y'|M_x$ is observed is positive (see Figure 5.2b). Consequently, we can still correctly infer unidirectional causality from X to Y.

However, as the coupling strength further increases, resulting in an increase in SL, we encounter difficulty in determining whether the causal link between X and Y is bidirectional or unidirectional, or we incorrectly infer bidirectional causality.

In summary, based on our analysis, we can accurately infer causality when the cou-



(a) CMS for cross map $X'|M_y$

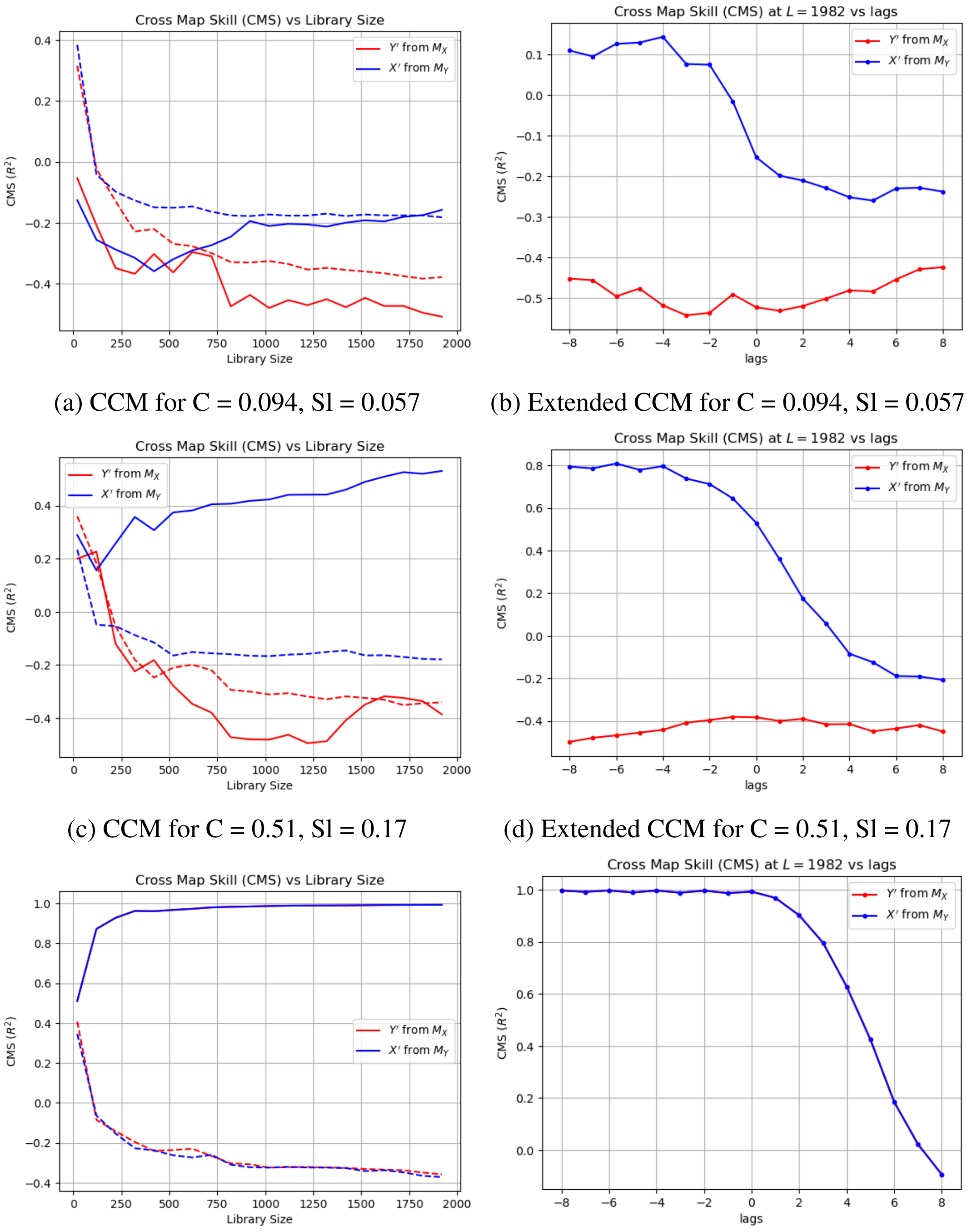


(b) CMS for cross-map $Y'|M_x$

Figure 5.2: Heat map with coupling strength and lag values as x and y axis, and maximum CMS represent by color (for identical henon system)

pling strength lies between 0.15 and 0.67, and the SL values range from 0.07 to 0.66.

In the figure5.3, we show representative CCM analysis for the three regions as defined in this section's introduction (5.1).



(a) CCM for C = 0.094, Sl = 0.057

(b) Extended CCM for C = 0.094, Sl = 0.057

(c) CCM for C = 0.51, Sl = 0.17

(d) Extended CCM for C = 0.51, Sl = 0.17

(e) CCM for C = 0.76, SL = 1.0

(f) Extended CCM for C = 0.76, SL = 1.0

Figure 5.3: CCM analysis for 3 (C, SL) pair, which are representative examples of CCM analysis in the 3 regions. [(a),(b)],[(c),(d)],[(e),(f)] are representative of Region I, Region II and Region III respectively

### 5.1.2 Case 2: Coupled non-identical Henon System

The system is described using the equation 4.1. The parameters are $a_1 = 1.4, b_1 = 0.3, b_2 = 0.2$. $C$ is the coupling strength and is varied from 0.01 to 1.2.

In this case, we also observe the trend where the Synchrony Likelihood (SL) increases with an increase in coupling strength, as shown in Figure 5.4. For very weak coupling strengths (C < 0.2), the Convergent Cross Mapping (CCM) approach is not able to identify the causal link between the variables. The SL values are less than 0.1 within this region. As the coupling strength increases, the CCM framework successfully identifies the causal links between the variables. This is observed until the coupling strength reaches approximately 0.8 and the SL value is around 0.6. However, when the coupling strength is further increased beyond this point, CCM methods begin to fail in accurately determining the causality between the variables. (refer figure (5.4,5.5a,5.5b).

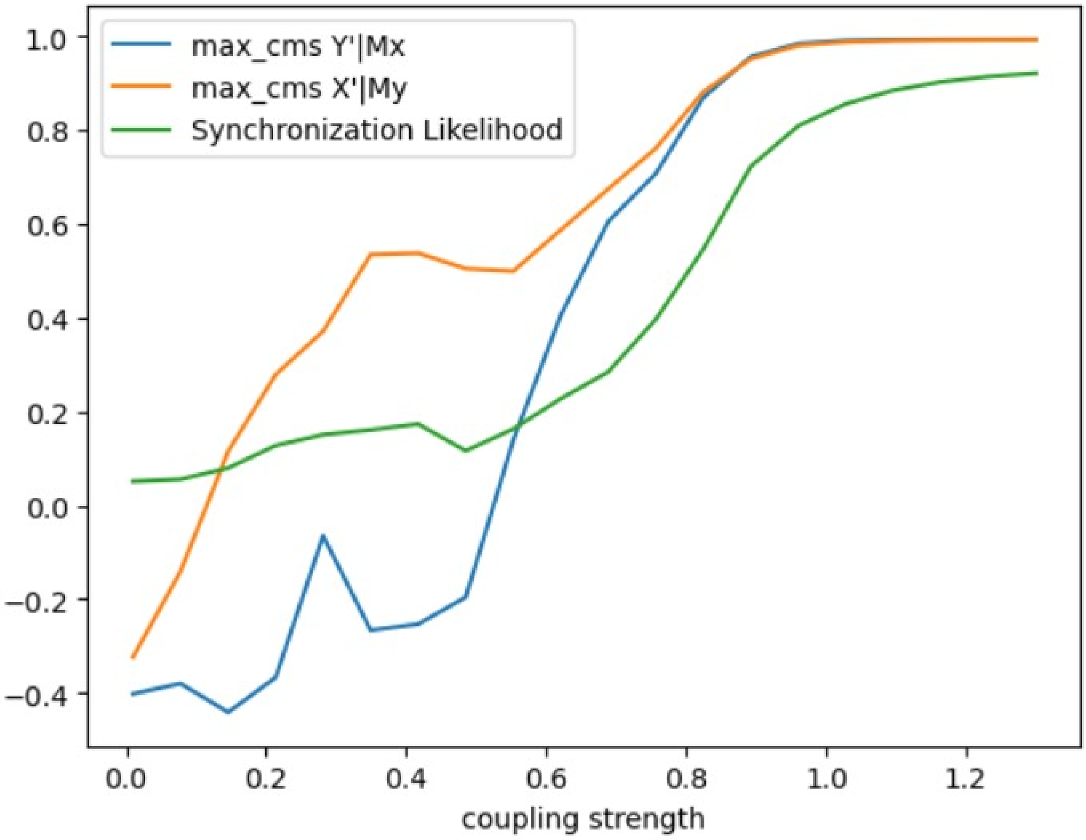
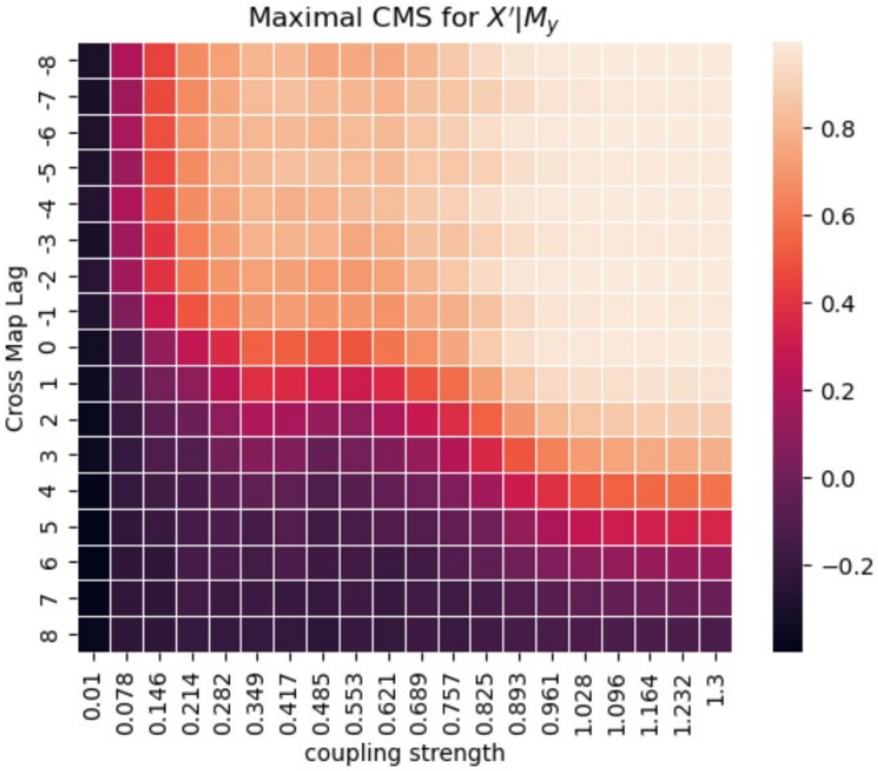
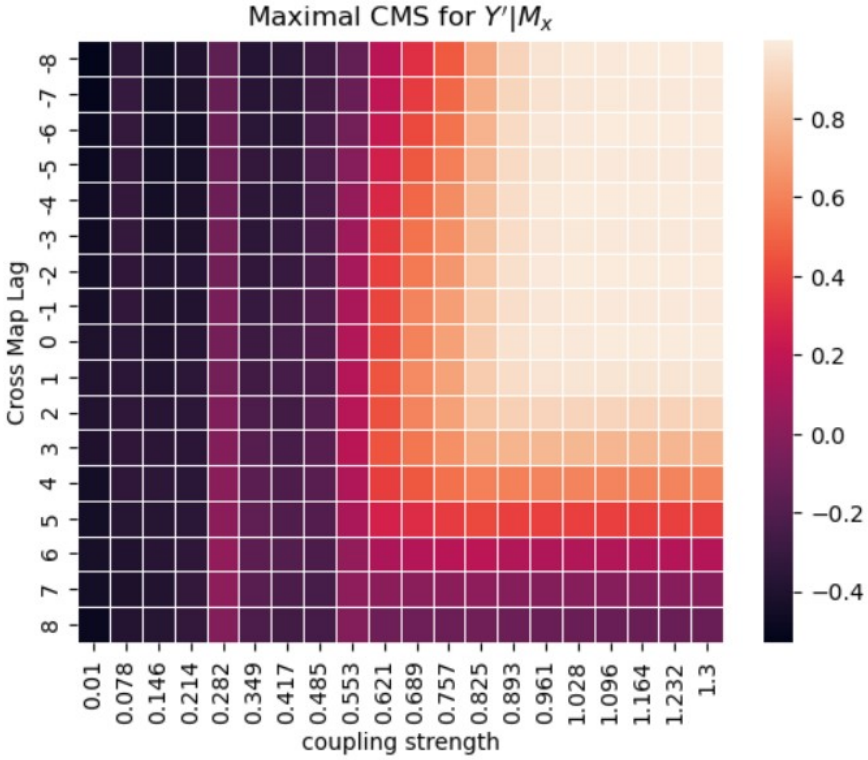


Figure 5.4: Plot of maximum CMS achieved for $X'|M_y and Y'|M_x$ with increasing coupling strength for non identical henon system. Sl vs coupling strength is also plotted



(a) CMS for cross map $X'|M_y$



(b) CMS for cross-map $Y'|M_x$

Figure 5.5: Heat map with coupling strength and lag values as x and y axis, and maximum CMS represent by color(for non identical henon system)

### 5.1.3 Case 3: Coupled 2D Logistic Map

Equation 4.2 defines 2D logistic system. The parameters values used for this analysis are $a_{11} = 3.8, a_{12} = 0, a_{22} = 3.3$. The parameter $a_{21}$ is the coupling parameter from X to Y and it is varied from 0.2 to 1 for this analysis. We will refer to $a_{12}$ as coupling strength.

In this case, we observe that increasing coupling strength doesn't necessarily lead to an increase in SL. As C increases, SL first increases and then falls (see figure 5.6). The SL > 0.1 for all values of C and SL< 0.6 for this case. We correctly infer causality for all the coupling strength values in consideration (see figures **??**).

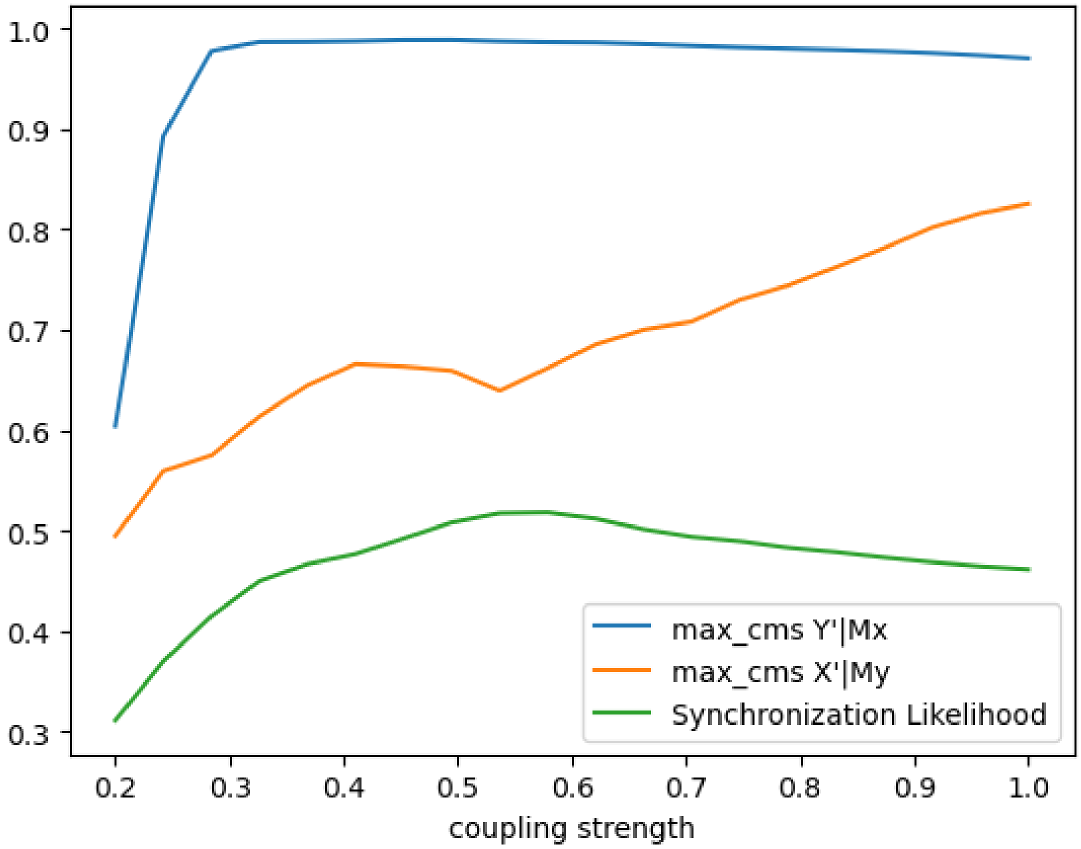
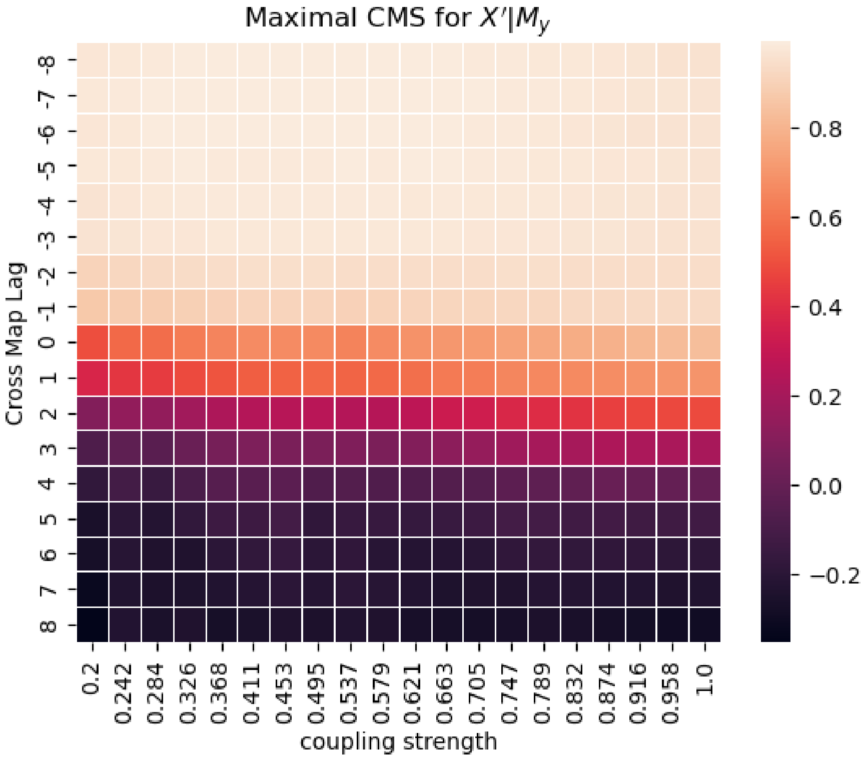
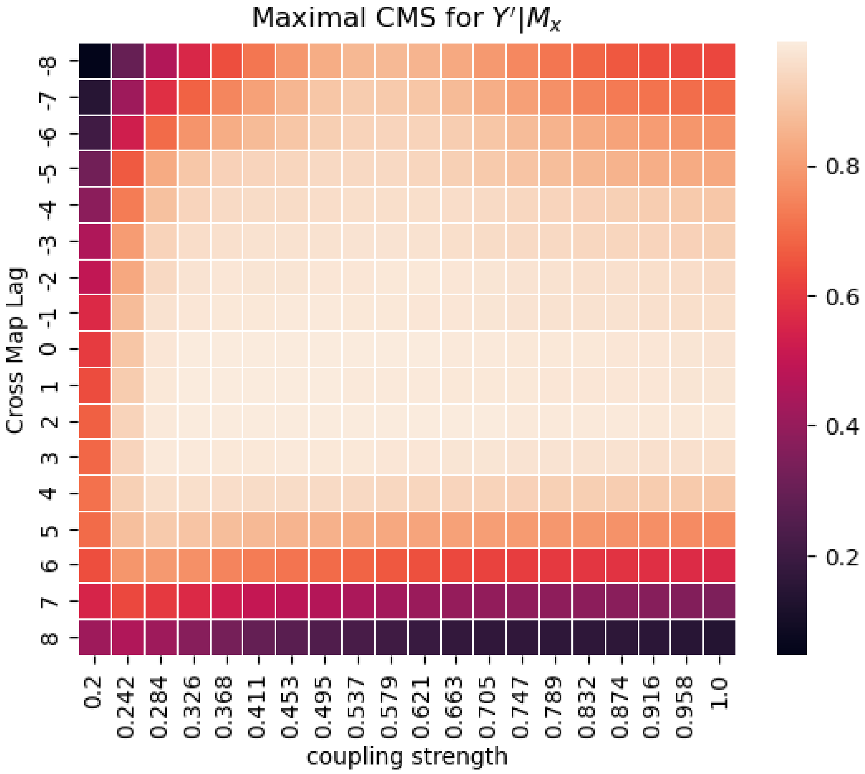


Figure 5.6: Plot of maximum CMS achieved for $X'|M_y and Y'|M_x$ with increasing coupling strength for coupled logistic map. Sl vs coupling strength is also plotted



(a) CMS for cross map $X'|M_y$



(b) CMS for cross-map $Y'|M_x$

Figure 5.7: Heat map with coupling strength and lag values as x and y axis, and maximum CMS represent by color(for coupled logistic map)

### 5.1.4 SL and Noise

For both identical and non-identical coupled Henon systems, we simulated SL for different coupling strengths and different SNRs. The results are shown in figure 5.8,5.9.

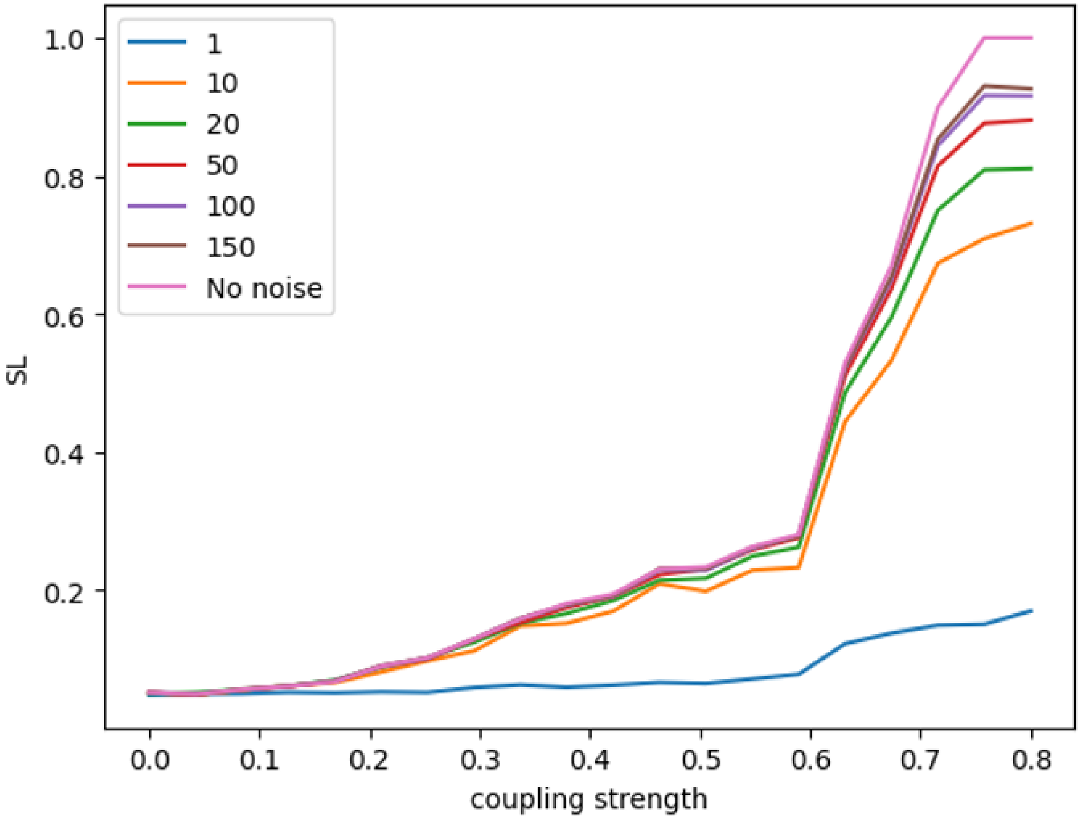


Figure 5.8: Sl values for various SNRs, identical coupled henon map

Lowering SNR leads to a decrease in Synchrony Likelihood (SL) values. However, it is important to note that the overall characteristics of the curve remain similar unless the noise levels become extremely high. The relationship between SNR and SL suggests that as the noise in the system increases, the synchronization between the time series becomes more challenging to detect accurately. This is reflected in the lower SL values observed under lower SNR conditions.

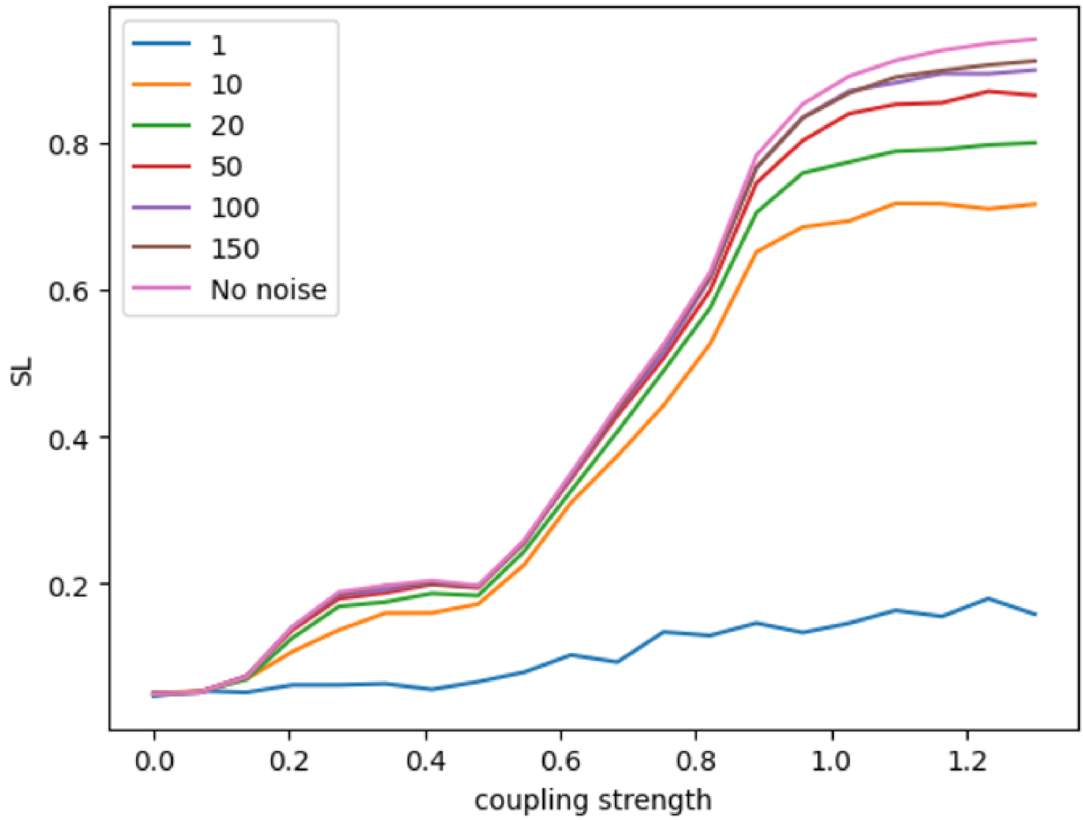


Figure 5.9: Sl values for various SNRs, non-identical coupled henon map

## 5.2 A failure mode for Multivariate CCM in differentiating direct from indirect link

Consider a 3D coupled logistic map given by equation 4.3 with parameters:

$$a_{11} = 3.9, a_{12} = 0.93, a_{13} = 0$$

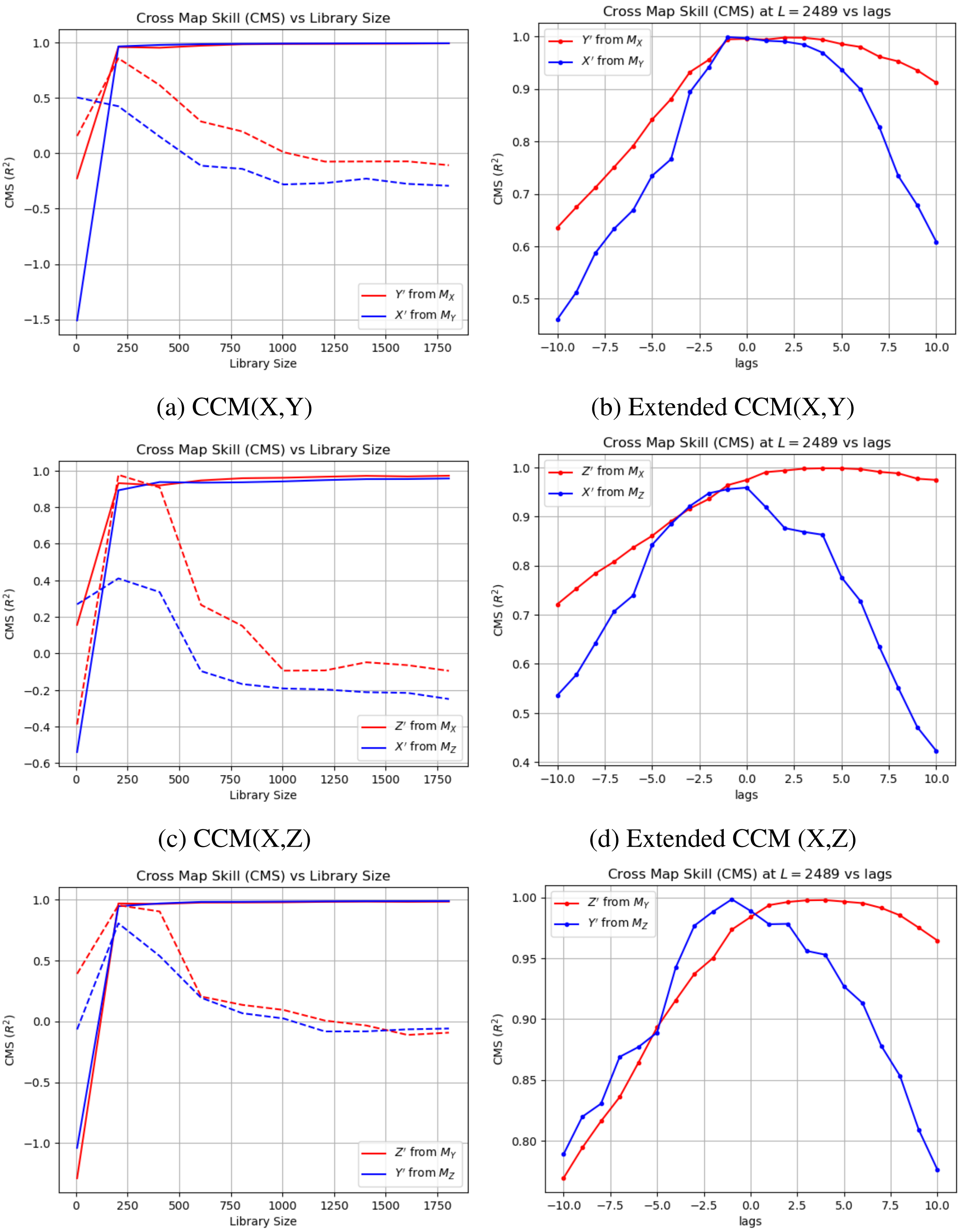
$$a_{21} = 0.8, a_{22} = 3.36, a_{23} = 0$$

$$a_{31} = 0, a_{32} = 0.45, a_{33} = 3.12$$

The true causal structure of this system is $X \leftrightarrow Y \rightarrow Z$ Figure 5.10 shows the result of pairwise CCM. The inferred causal links between variables are $X \leftrightarrow Y$, $Y \rightarrow Z$, $X \rightarrow Z$.

To identify whether $X \leftarrow Z$ is a direct link or indirect we perform multivariate CCM. The result is shown in figure 5.11

In Figure 5.11a we don't see a significant decrease in CMS for any of the effect variables Y and Z compared to the multivariate shadow manifold and we wrongly conclude



(a) CCM(X,Y)

(b) Extended CCM(X,Y)

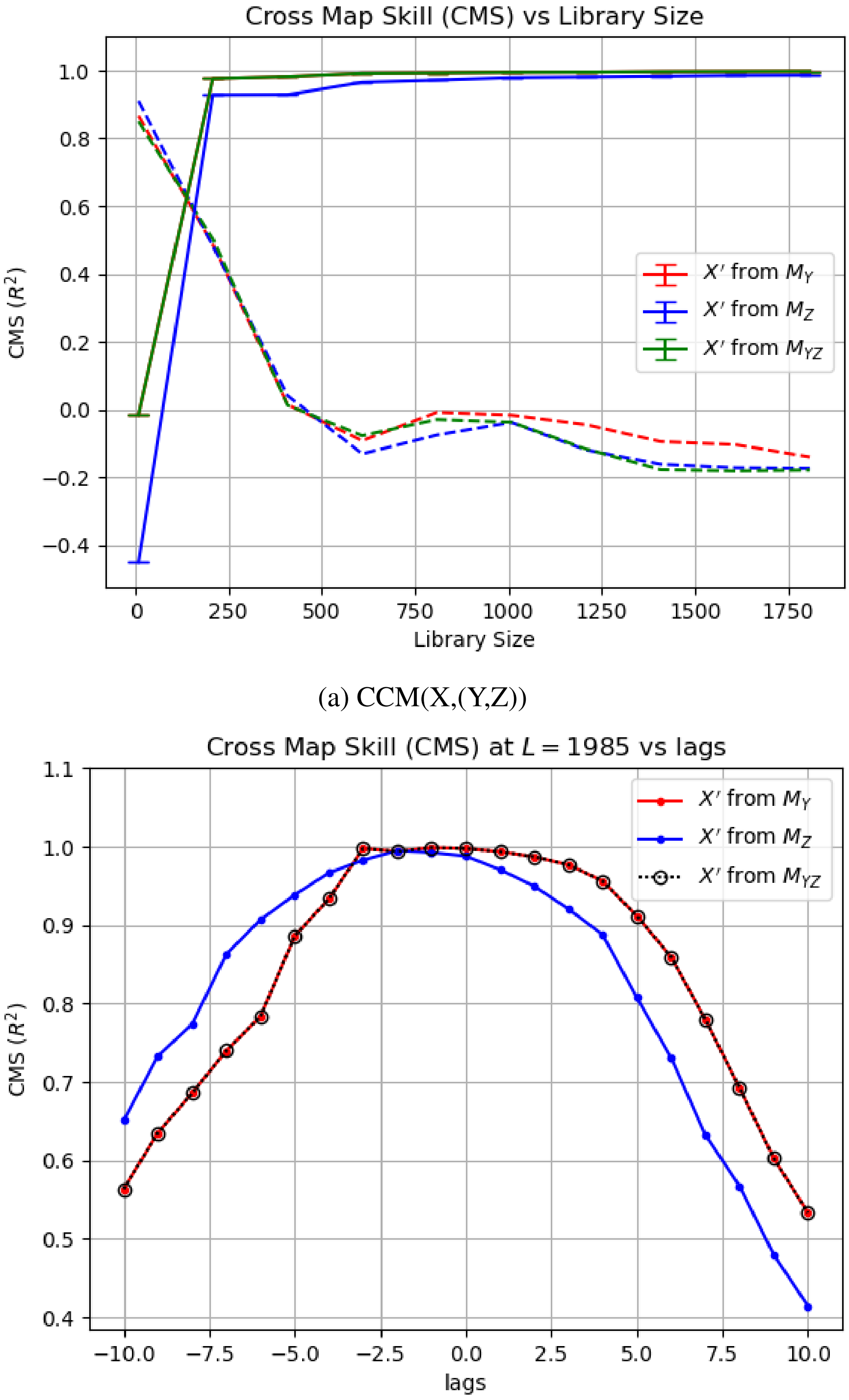
(c) CCM(X,Z)

(d) Extended CCM (X,Z)

(e) CCM(Y,Z)

(f) Extended CCM(Y,Z)

Figure 5.10: Pairwise CCM Analysis



(a) CCM(X,(Y,Z))

(b) Extended CCM(X,(Y,Z))

Figure 5.11: Multivariate CCM Analysis)

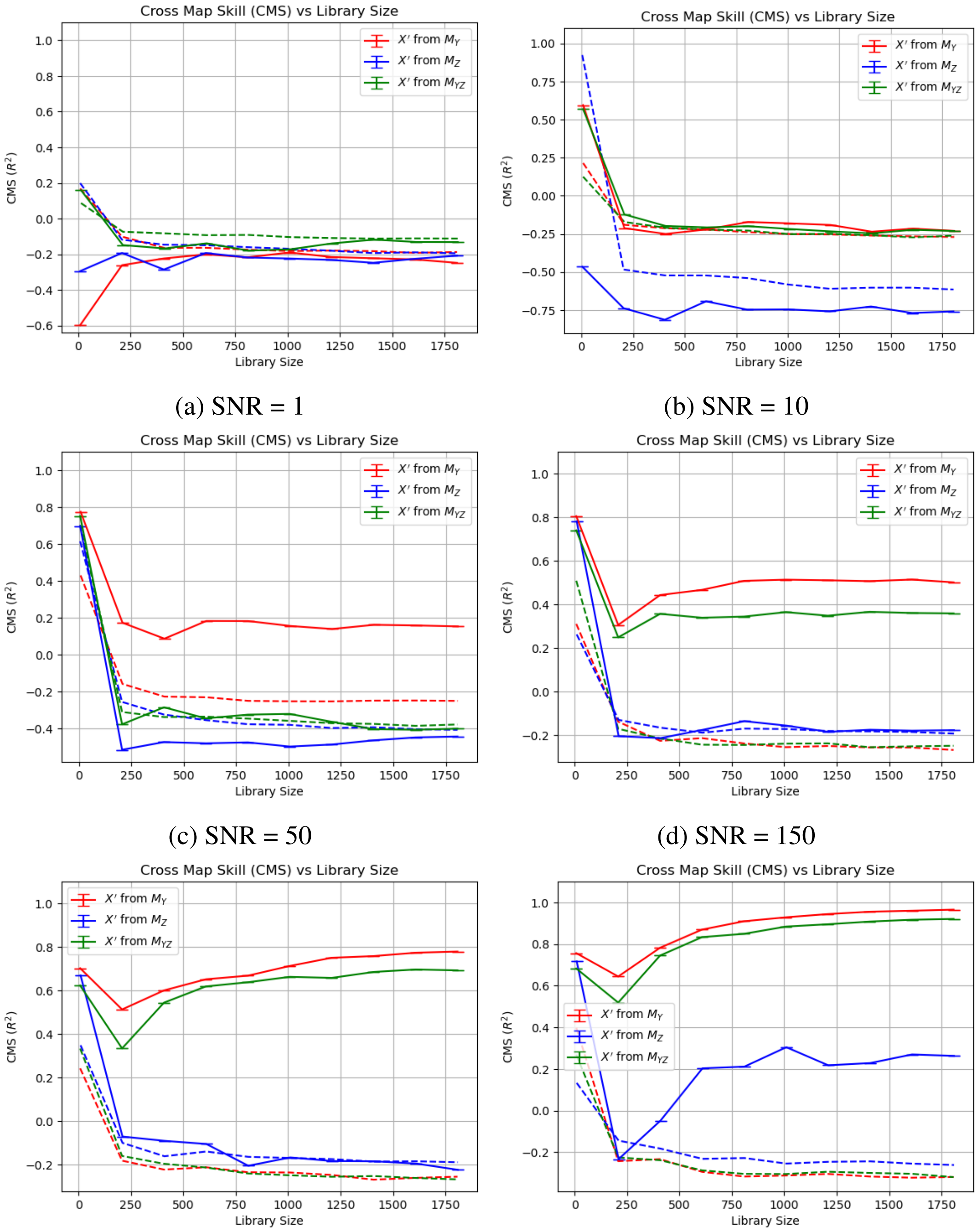
that the identified causal links are direct.

### 5.3 Noise and Multivariate CCM

For a 3D coupled logistic map given by equation 4.3 with parameters:

$$a_{11} = 3.9, a_{12} = 0, a_{13} = 0$$
$$a_{21} = 0.15, a_{22} = 3.3, a_{23} = 0$$
$$a_{31} = 0, a_{32} = 0.12, a_{33} = 3.5$$

The true causal structure is given by $X \rightarrow Y \rightarrow Z$.For the given equation, we conducted simulations of multivariate CCM (X, (Y, Z)) considering measurement noise at various signal-to-noise ratios (SNRs). The results of these simulations are presented in figure 5.12,5.13.



(a) SNR = 1

(b) SNR = 10

(c) SNR = 50

(d) SNR = 150

(e) SNR = 500

(f) No noise

Figure 5.12: Multivariate CCM(X,(Y,Z)) for multiple SNRs

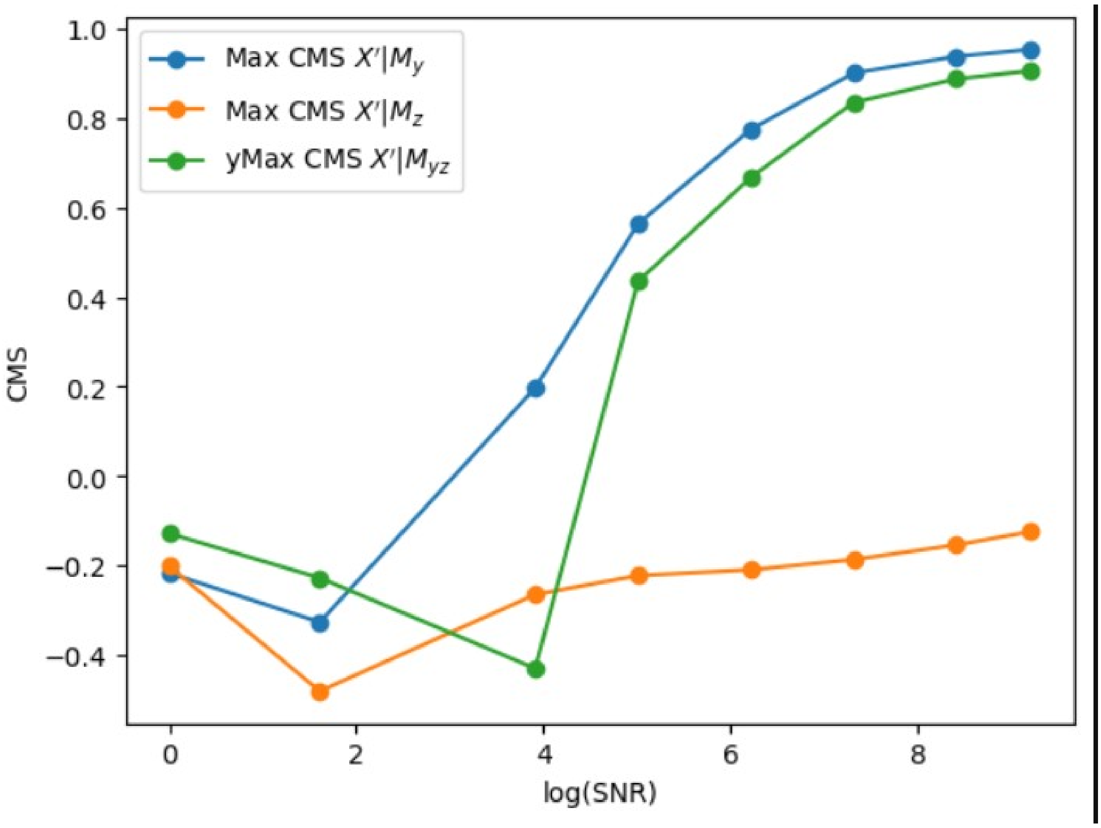


Figure 5.13: CMS at maximum library length vs Log(SNR)

# CHAPTER 6

# DISCUSSION

## 6.1 Synchronization Likelihood as the Metric to gauge applicability of CCM

Sugihara *et al.* (2012) points out that CCM may not work for a system with strong unidirectional forcing that leads to general synchrony. While the statement implies a direct relationship between strong unidirectional forcing and synchrony, we have seen this relationship is not necessarily the case. In fact, it is possible that a higher level of synchrony can be achieved with lower coupling strength, and an increase in coupling strength may actually lead to a reduction in the level of general synchrony (GS) as is the case for 2D coupled logistic maps. Based on this work, we propose an alternative hypothesis that the applicability of the CCM framework is limited not by the coupling strength itself, but rather by the level of general synchrony between the systems. We attribute this phenomenon to the use of the Simplex projection-based method for recovering cause values. When recovering, for example, variable X from $M_y$, we identify the nearest neighbors of the embedding vector $\bar{y}(t)$ and determine the contemporaneous value $x(t)$ based on these neighbors. The recovered value is then calculated as a weighted mean. As the level of general synchrony increases, it becomes more likely that the neighborhood of $M_x$ corresponds to the neighborhood of $M_y$, as illustrated in Figure 6.1. When the points are in close proximity within the attractor of the forcing variable, an increase in synchrony causes the contemporaneous points in the forced system to also come closer. We have chosen Synchronization likelihood (SL) as our measure of synchrony due to its state-based nature and the advantage of being a normalized measure.

In our experimental investigations, we have observed that the cutoff for synchronization likelihood (SL) values, above which the Convergent Cross Mapping (CCM) framework fails, is not precisely defined and varies depending on the specific system being studied. For practitioners, we suggest a heuristic based on our simulations: when SL values are below 0.6, it is more likely that the CCM framework will accurately infer the causal structure.

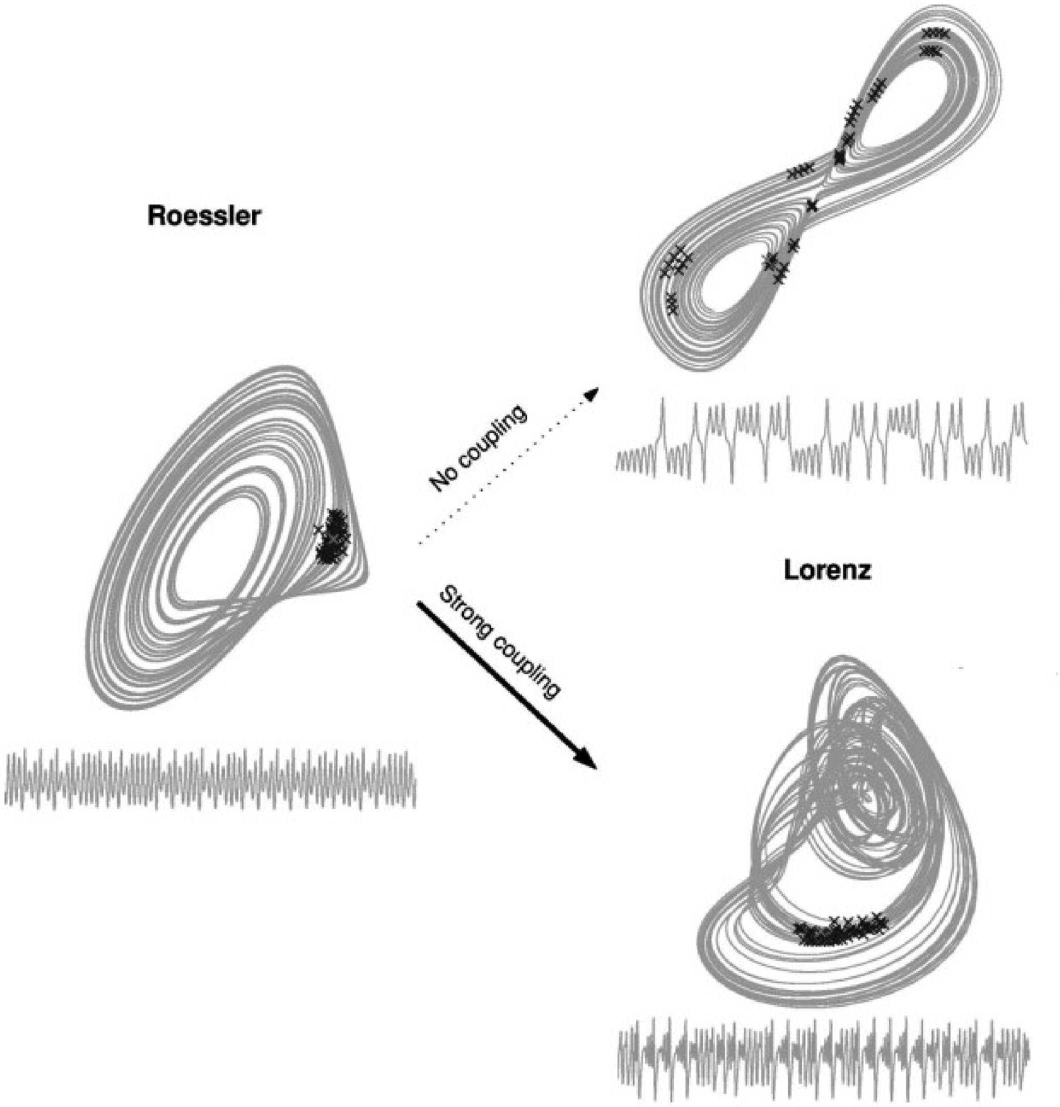


Figure 6.1: Figure showing how the attractor of forced system changes under strong forcing that causes synchronization. Figure taken from Quiroga *et al.* (2002)

The variability in the SL cutoff can be likely attributed to the limitations imposed by the use of limited time series data for performing CCM and calculating SL. This inherent constraint inadvertently introduces approximations into the analysis.

### 6.2 Failure Mode of Multivariate CCM

In addition to the findings mentioned earlier, this research also highlights a specific class of systems in which multivariate Convergent Cross Mapping (CCM) exhibits poor performance. These systems are characterized by the presence of a causal structure in the form of $X \leftrightarrow Y \rightarrow Z$, where there is a bidirectional causal link between X and Y, and Y causally influences Z. Importantly, the causal link between X and Y demonstrates a high level of synchrony.

In the case presented in Section 5.2, the synchronization likelihood (SL) values between X and Y are approximately 0.5. This scenario represents a failure mode for multivariate CCM because the coupling between X and Y exhibits a high level of synchrony. Consequently, the recovery of X and Y from the manifold of Z is similarly accurate. In the absence of noise in the time series, this recovery is nearly perfect.

### 6.3 Effects of Noise

In our observations of synchronization likelihood (SL) values, we have noticed a decrease in SL as noise levels increase. It is important to note that while the SL values decrease, the overall characteristics of the SL curve remain relatively consistent, unless the noise levels reach extremes. This finding may help explain the effectiveness of the controlled noise injection method proposed by Mønster *et al.* (2017), as lower SL values tend to result in more accurate CCM analysis.

Furthermore, we have explored the impact of noise on multivariate CCM. Although further extensive research is necessary to draw generalized conclusions, it is interesting to note, perhaps unsurprisingly, that the slope of the maximum CMS versus log(SNR) is smallest for $X'|M_z$, which represents the recovery of X from the effects it indirectly causes via Y (Figure 5.13). The information about X present in Z is mediated through Y, implying that it contains less direct information about X. Consequently, a smaller amount of noise is capable of disrupting the recovery of this indirect information.

# CHAPTER 7

# CONCLUSION AND FUTURE WORK

CCM is a promising tool for identifying causal structures in nonlinear dynamical systems, but there are still limitations to its practical application, and further research is needed to deepen our understanding of the underlying theory. In this work, we have attempted to address some of these challenges.

One significant contribution of this work is the proposal of an alternative hypothesis, suggesting that the applicability of the CCM framework is not limited by coupling strength but by the level of general synchrony. We explored the use of synchronization likelihood (SL) as a measure to quantify the level of general synchrony and as a metric to distinguish datasets suitable for reliable CCM analysis. This approach provides a means of assessing the feasibility of CCM.

Additionally, we identified a failure mode for multivariate CCM, shedding light on its limitations in systems characterized by strong synchrony between causally connected variables. Understanding these failure modes is essential for refining and improving the performance of the CCM framework. Furthermore, we also investigated the effects of noise on the CCM framework and its implications. The impact of noise on the CCM framework and nonlinear systems at large poses a significant challenge.

Future work should aim to further investigate and mathematically explore the proposed alternative hypothesis, potentially utilizing topological tools. Exploring how process noise affects the performance and reliability of CCM is a significant area for further investigation. Moreover, delving deeper into the limitations and applicability of multivariate CCM is an essential avenue for future research. Understanding the specific conditions and scenarios where multivariate CCM excels or struggles can contribute to its more effective and reliable application in practical settings.

# REFERENCES

1. **Bartsev, S.**, **M. Saltykov**, **P. Belolipetsky**, and **A. Pianykh** (2021). Imperfection of the convergent cross-mapping method. *IOP Conference Series: Materials Science and Engineering*, **1047**(1), 012081. URL `https://doi.org/10.1088/1757-899x/1047/1/012081`.

2. **Cabañero-Gomez, L.**, **R. Hervas**, **I. Gonzalez**, and **L. Rodriguez-Benitez** (2021). eeglib: A python module for EEG feature extraction. *SoftwareX*, **15**, 100745. URL `https://doi.org/10.1016%2Fj.softx.2021.100745`.

3. **Cummins, B.**, **T. Gedeon**, and **K. Spendlove** (2015). On the efficacy of state space reconstruction methods in determining causality. *SIAM Journal on Applied Dynamical Systems*, **14**(1), 335–381. URL `https://doi.org/10.1137/130946344`.

4. **Demiralp, S.** and **K. D. Hoover** (2003). Searching for the causal structure of a vector autoregression. *Oxford Bulletin of Economics and Statistics*, **65**(s1), 745–767. URL `https://doi.org/10.1046/j.0305-9049.2003.00087.x`.

5. **Deyle, E. R.** and **G. Sugihara** (2011). Generalized theorems for nonlinear state space reconstruction. *PLoS ONE*, **6**(3), e18295. URL `https://doi.org/10.1371/journal.pone.0018295`.

6. **Díaz, E.**, **J. E. Adsuara**, **Á. M. Martínez**, **M. Piles**, and **G. Camps-Valls** (2022). Inferring causal relations from observational long-term carbon and water fluxes records. *Scientific Reports*, **12**(1). URL `https://doi.org/10.1038/s41598-022-05377-7`.

7. **Feng, G.**, **J. G. Quirk**, and **P. M. Djuric**, Discovering causalities from cardiotocography signals using improved convergent cross mapping with gaussian processes. *In ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020*a*. URL `https://doi.org/10.1109/icassp40776.2020.9053462`.

8. **Feng, G.**, **K. Yu**, **Y. Wang**, **Y. Yuan**, and **P. M. Djurić**, Improving convergent cross mapping for causal discovery with gaussian processes. *In ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2020*b*.

9. **Glymour, C.**, **K. Zhang**, and **P. Spirtes** (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, **10**. URL `https://doi.org/10.3389/fgene.2019.00524`.

10. **Granger, C. W. J.** (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**(3), 424–438. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/1912791`.

11. **Harnack, D.**, **E. Laminski**, **M. Schünemann**, and **K. R. Pawelzik** (2017). Topological causality in dynamical systems. *Physical Review Letters*, **119**(9). URL `https://doi.org/10.1103/physrevlett.119.098301`.

12. **hui Lang, S.**, **H. Zhu**, **G. dong Sun**, **Y. Jiang**, and **C. ling Wei** (2021). A study on methods for determining phase space reconstruction parameters. *Journal of Computational and Nonlinear Dynamics*, **17**(1). URL `https://doi.org/10.1115/1.4052721`.

13. **Keskin, Z.** and **T. Aste** (2020). Information-theoretic measures for nonlinear causality detection: application to social media sentiment and cryptocurrency prices. *Royal Society Open Science*, **7**(9), 200863. URL `https://doi.org/10.1098/rsos.200863`.

14. **Khanmohammadi, S.** (2017). An improved synchronization likelihood method for quantifying neuronal synchrony. *Computers in Biology and Medicine*, **91**, 80–95. URL `https://doi.org/10.1016%2Fj.compbiomed.2017.09.022`.

15. **Krakovská, A.**, **Š. Pócoš**, **K. Mojžišová**, **I. Bečková**, and **J. X. Gubáš** (2022). State space reconstruction techniques and the accuracy of prediction. *Communications in Nonlinear Science and Numerical Simulation*, **111**, 106422. URL `https://doi.org/10.1016/j.cnsns.2022.106422`.

16. **Ma, H.**, **K. Aihara**, and **L. Chen** (2014). Detecting causality from nonlinear dynamics with short-term time series. *Scientific Reports*, **4**(1). URL `https://doi.org/10.1038/srep07464`.

17. **Malinsky, D.** and **P. Spirtes**, Causal structure learning from multivariate time series in settings with unmeasured confounding. *In* **T. D. Le**, **K. Zhang**, **E. Kıcıman**, **A. Hyvärinen**, and **L. Liu** (eds.), *Proceedings of 2018 ACM SIGKDD Workshop on Causal Disocvery*, volume 92 of *Proceedings of Machine Learning Research*. PMLR, 2018. URL `https://proceedings.mlr.press/v92/malinsky18a.html`.

18. **McCracken, J. M.** and **R. S. Weigel** (2014). Convergent cross-mapping and pairwise asymmetric inference. *Physical Review E*, **90**(6). URL `https://doi.org/10.1103%2Fphysreve.90.062903`.

19. **Montez, T.**, **K. Linkenkaer-Hansen**, **B. van Dijk**, and **C. Stam** (2006). Synchronization likelihood with explicit time-frequency priors. *NeuroImage*, **33**(4), 1117–1125. URL `https://doi.org/10.1016%2Fj.neuroimage.2006.06.066`.

20. **Mønster, D.**, **R. Fusaroli**, **K. Tylén**, **A. Roepstorff**, and **J. F. Sherson** (2017). Causal inference from noisy time-series data — testing the convergent cross-mapping algorithm in the presence of noise and external influence. *Future Generation Computer Systems*, **73**, 52–62. ISSN 0167-739X. URL `https://www.sciencedirect.com/science/article/pii/S0167739X16307427`.

21. **Nithya, S.** and **A. K. Tangirala**, Multivariable causal analysis of nonlinear dynamical systems using convergent cross mapping. *In 2021 Seventh Indian Control Conference (ICC)*. 2021.

22. **Quiroga, R. Q.**, **A. Kraskov**, **T. Kreuz**, and **P. Grassberger** (2002). Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Physical Review E*, **65**(4). URL `https://doi.org/10.1103/physreve.65.041903`.

23. **Rulkov, N. F.**, **M. M. Sushchik**, **L. S. Tsimring**, and **H. D. I. Abarbanel** (1995). Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E*, **51**(2), 980–994. URL https://doi.org/10.1103%2Fphysreve.51.980.

24. **Runge, J.**, **P. Nowack**, **M. Kretschmer**, **S. Flaxman**, and **D. Sejdinovic** (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, **5**(11). URL https://doi.org/10.1126/sciadv.aau4996.

25. **Sauer, T.**, **J. A. Yorke**, and **M. Casdagli** (1991). Embedology. *Journal of Statistical Physics*, **65**(3-4), 579–616. URL https://doi.org/10.1007%2Fbf01053745.

26. **Stam, C.** and **B. van Dijk** (2002). Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. *Physica D: Nonlinear Phenomena*, **163**(3), 236–251. ISSN 0167-2789. URL https://www.sciencedirect.com/science/article/pii/S0167278901003864.

27. **Stark, J.** (1999). Delay embeddings for forced systems. i. deterministic forcing. *Journal of Nonlinear Science*, **9**(3), 255–332. URL https://doi.org/10.1007/s003329900072.

28. **Stark, J.**, **D. Broomhead**, **M. Davies**, and **J. Huke** (2003). Delay embeddings for forced systems. II. stochastic forcing. *Journal of Nonlinear Science*, **13**(6), 519–577. URL https://doi.org/10.1007/s00332-003-0534-4.

29. **Strogatz, S. H.**, *Nonlinear Dynamics and Chaos*. CRC Press, 2018. URL https://doi.org/10.1201/9780429492563.

30. **Sugihara, G.**, **R. May**, **H. Ye**, **C. hao Hsieh**, **E. Deyle**, **M. Fogarty**, and **S. Munch** (2012). Detecting causality in complex ecosystems. *Science*, **338**(6106), 496–500. URL https://www.science.org/doi/abs/10.1126/science.1227079.

31. **Takens, F.**, Detecting strange attractors in turbulence. *In Lecture Notes in Mathematics*. Springer Berlin Heidelberg, 1981, 366–381. URL https://doi.org/10.1007%2Fbfb0091924.

32. **Vlachos, I.** and **D. Kugiumtzis**, State space reconstruction from multiple time series. *In Topics on Chaotic Systems*. WORLD SCIENTIFIC, 2009. URL https://doi.org/10.1142/9789814271349_0043.

33. **Ye, H.**, **E. R. Deyle**, **L. J. Gilarranz**, and **G. Sugihara** (2015). Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific Reports*, **5**(1). URL https://doi.org/10.1038/srep14750.

34. **Yuan, A. E.** and **W. Shou** (2022). Data-driven causal analysis of observational biological time series. *eLife*, **11**. URL https://doi.org/10.7554/elife.72518.