Causality in Time Series Systems

Master's Thesis at the Faculty of Physics
of the
Ludwig-Maximilian-University Munich

submitted by
Haochun Ma
born in Augsburg on the 27th May, 1997

Munich, Germany, 7th September, 2020

Supervisor:
PD Dr. Christoph Räth

Kausalität in Zeitreihen Systemen

Masterarbeit an der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von
Haochun Ma
geboren in Augsburg am 27. Mai 1997

München, den 7. September 2020

Betreuer:
PD Dr. Christoph Räth

Abstract

Causality inference for time series systems has been subject to intensive research across many generations of physicist and, in light of the boom of computational resources, has been increasingly applied to a wider range of areas such as biology or finance. In this thesis we structurally compare three inference methods, *Granger Causality*, *Transfer Entropy*, and *Convergent Cross Mapping*, by applying them to synthetic nonlinear systems. While we verify that Granger Causality only detects linear causal relations, our analysis with *Fourier Transform* surrogates shows that a significant amount of causality, measured by Transfer Entropy and Convergent Cross Mapping, is driven by nonlinear properties. Our study of the Lorenz attractor further suggests different structures for different timeframe lengths. Upon introducing measures for the system causality, we observe that the long-term causality of the system remains approximately constant with a major nonlinear component. On a short-term scale, the causality resolution changes, which we can map to certain locations within the attractor. We find these properties to apply to several other synthetic nonlinear systems. The resulting framework is designed to be applicable to real-world time series systems in order to detect unknown causality structures and drivers in other research areas.

Contents

Chapter 1

Introduction

Causality, as one of the fundamental principles of physics, has been an area of intensive research across all generations of scientists and hence many different concepts evolved arising from the evergrowing endeavor and complexity of physical theories.

While in the classical understanding of Newton *actio* and *reactio* were defined to be simultaneously coupled, Einstein introduced a temporal and spatial component by defining causality as events connected through the light cone of general relativity. Subsequently, the disruption of quantum mechanics led to a probability-dominated understanding of physics with causality being an inconceivable concept in a non-deterministic world. With the emergence of chaos theory, causality was put into the context of stability and equilibria of dynamical systems, which became known to the general public as the butterfly effect.

Analogously to the definitions of causality, inference methods also evolved and increased in complexity in the course of time. Especially the explosion of computational resources led to the development of algorithms designed to infer causality within time series systems. These statistical models have been applied across various fields ranging from biology to finance.

The evolution of causality inference took a similar path as causality itself, with Granger introducing a model capturing similar patterns using regression of lagged time series. This idea agrees with our common understanding of causality in a temporal shifted order of events. The major drawback of *Granger Causality*, which measures only linear dependencies, was addressed by *Transfer Entropy*. Using probability-dependent measures from information theory, it compares the amount of uncertainty reduced between two time series. Following the studies of chaotic nonlinear systems such as the famous Lorenz attractor, causality inference methods started building on the reconstruction of the dynamical structure between coupled variables. One of the most recent breakthroughs within these state space reconstruction methods is

Convergent Cross Mapping, which is based on the transitive relation present in the topology of dynamical systems.

These three structurally different techniques form the main research question of this thesis, in which we analyze the origin and drivers of causality in synthetic nonlinear dynamic systems. Thus, we methodically compare the inference algorithms before we detect causality drivers using *Fourier transform* surrogate data. This sets the foundation for applications to real-world time series systems.

Therefore, we begin by introducing the data sources and preprocessing techniques in chapter 2. Subsequently, we set up the mathematical framework and methodologies to analyze time series systems in Chapter 3. Before we perform our causality driver analysis in Chapter 5, we describe the respective inference algorithms in Chapter 4.

Chapter 2

Data

This chapter gives an overview of the underlying data sources and different preprocessing techniques. A special emphasis is placed on the calculation of time series surrogates. Before we dive deeper into the specific nature of the nonlinear systems analyzed in this thesis, we briefly define the general concept of a time series system and establish a mathematical notation.

2.1 Time Series Systems

We understand a time series system $\mathcal{X}$ as a sequence of measurements $\mathcal{M}$ of a dynamical system $\mathcal{S}$, which is in essence a map between time and a time-dependent state space Ω_t:

$$\begin{aligned} \mathcal{S} : \mathbb{R} &\longrightarrow \Omega_t \\ t &\longmapsto \omega \,. \end{aligned}$$

Since we generally can and only want to observe a finite subset of the state space $\tilde{\Omega}_t$, we define a single measurement of N state variables $\{X_1, \ldots, X_N\}$ at time t as follows:

$$\begin{aligned} \mathcal{M} : \mathbb{R} &\longrightarrow \tilde{\Omega}_t \subseteq \mathbb{R}^N \cap \Omega_t \\ t &\longmapsto x_n \,, \end{aligned}$$

with $n \in \{1, \ldots, N\}$ linking the measurement x_n to its corresponding state variable X_n.

We perform a number of T measurements at an ascending sequence of time points $\boldsymbol{t} = (t_1, \ldots, t_T)$:

$$\begin{aligned} \mathcal{M}^{(T)} : \mathbb{R}^T &\longrightarrow \tilde{\Omega}_{t_1} \times \ldots \times \tilde{\Omega}_{t_N} \subseteq \mathbb{R}^{N \times T} \cap (\Omega_{t_1} \times \ldots \times \Omega_{t_N}) \\ \boldsymbol{t} &\longmapsto x_{n,t} \equiv \{\mathcal{M}(\boldsymbol{t}_t) \mid t \in \{1, \ldots, T\}\}, \end{aligned} \tag{2.1}$$

where t denotes the t-th element of the sequence $\boldsymbol{t}$.

It is important to note that according to this shift in definition $t \in \mathbb{N}$ henceforth denotes the index of the time point (or timestep) rather than the time point itself.

Hence, the tuple $\mathcal{X} \equiv (\boldsymbol{t}, x_{n,t})$ defines a *complete* time series system. However, for simplicity reasons we will in most cases only specify $x_{n,t}$.

The system is called N-dimensional with time series length T for $n \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$.

While this formulation may seem counterintuitive, it allows great flexibility using the following index notation:

- If both n and t are specified as numbers, then $x_{n,t}$ denotes the n-th measured state space variable X_n at the t-th timestep:

$$x_{n,t} \in \mathbb{R}.$$

- If n is specified as a number and $t \in \{1, \ldots, T\}$ as a set, then $x_{n,\underline{T}}$ denotes the time series of X_n:

$$x_{n,\underline{T}} \equiv (x_{n,1}, x_{n,2}, \ldots, x_{n,T}) \in \mathbb{R}^T.$$

- If $n \in \{1, \ldots, N\}$ is specified as a set and t as a number, then $x_{\underline{N},t}$ denotes the measured state space at the t-th timestep:

$$x_{\underline{N},t} \equiv \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{N,t} \end{pmatrix} \in \mathbb{R}^N.$$

- If both $n \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$ are specified as sets, then $x_{\underline{N},\underline{T}}$ denotes the time series system as a matrix with rows being time series and columns being the measured state space variables at each timestep:

$$x_{\underline{N},\underline{T}} \equiv \begin{pmatrix} x_{1,\underline{T}} \\ x_{2,\underline{T}} \\ \vdots \\ x_{N,\underline{T}} \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,T} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,T} \end{pmatrix} \in \mathbb{R}^{N \times T}.$$

Furthermore, for $n \in \{n_0, \ldots, n_e\}$ and $t \in \{t_0, \ldots, t_e\}$ with $n_0 \geq 1$, $t_0 \geq 1$, $n_e \leq N$, and $t_e \leq T$ we establish shifted indices. The subsystem of $x_{\underline{N},\underline{T}}$ corresponding to the indices is denoted as:

$$
\begin{aligned}
x_{n_0:n_e,t_0:t_e} &\equiv \begin{pmatrix} x_{n_0,t_0:t_e} \\ x_{n_0+1,t_0:t_e} \\ \vdots \\ x_{n_e,t_0:t_e} \end{pmatrix} = \left(x_{n_0:n_e,t_0}, x_{n_0:n_e,t_0+1,\ldots,x_{n_0:n_e,t_e}}\right) \\
&= \begin{pmatrix} x_{n_0,t_0} & x_{n_0+1,t_0+1} & \cdots & x_{n_0,t_e} \\ x_{n_0+1,t_0} & x_{n_0+1,t_0+1} & \cdots & x_{n_0+1,t_e} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_e,t_0} & x_{n_e,t_0+1} & \cdots & x_{N,t_e} \end{pmatrix} \in \mathbb{R}^{(n_e-n_0+1)\times(t_e-t_0+1)} .
\end{aligned}
$$

This notation is comparable to the vector notation used in several programming languages, e.g. MATLAB or Python. Generally, the structure of this thesis is designed to enable an easy implementation of the model.

2.2 Synthetic Systems

While the true causal nature in real-world systems is usually unknown, synthetic systems exhibit causal behavior according to the interaction parameters between their state variables. This allows us to verify the validity of our methods by analyzing synthetic systems with different types and strengths of causality. We obtain the time series system data through solving the defining equations of the systems, which are in our case recurrence relations and *Ordinary Differential Equations* (ODEs). Furthermore, we create simple synthetic control systems, which serve as dummies for our model.

Note that in the following we comply to the conventional notation of the systems and disregard the technicalities established in Section 2.1.

2.2.1 Coupled Logistic

In order to reproduce the *Convergent Cross Mapping* (CCM) causality inference technique, we reproduce the coupled two-species nonlinear logistic system specified by Sugihara et al. [1]:

$$\begin{aligned} x_{t+1} &= x_t \cdot [r_x - r_x \cdot x_t - \beta_{y,x} \cdot y_t] \\ y_{t+1} &= y_t \cdot [r_y - r_y \cdot y_t - \beta_{x,y} \cdot x_t] \,, \end{aligned} \tag{2.2}$$

where $r_x = 3.8$, $r_y = 3.5$, $\beta_{x,y} = 0.1$, and $\beta_{y,x} = 0.02$ are set as the default parameters. Note that we switched the indices of the couplings $\beta_{x,y}$ and $\beta_{y,x}$, so that the variable in the first index drives the second. We use the initial condition $(x_1, y_1) = (0.2, 0.4)$ for our simulation.

The major benefit of this system is the nonlinear and chaotic behavior it exhibits despite of its simpleness. As Sugihara et al. [1] described, it switches between phases of anti-correlation, coherence, and randomness. This property, called mirage correlation, is illustrated in Figure 2.1 where the system starts with an anti-correlated phase, followed by a period of coherence, and so forth.

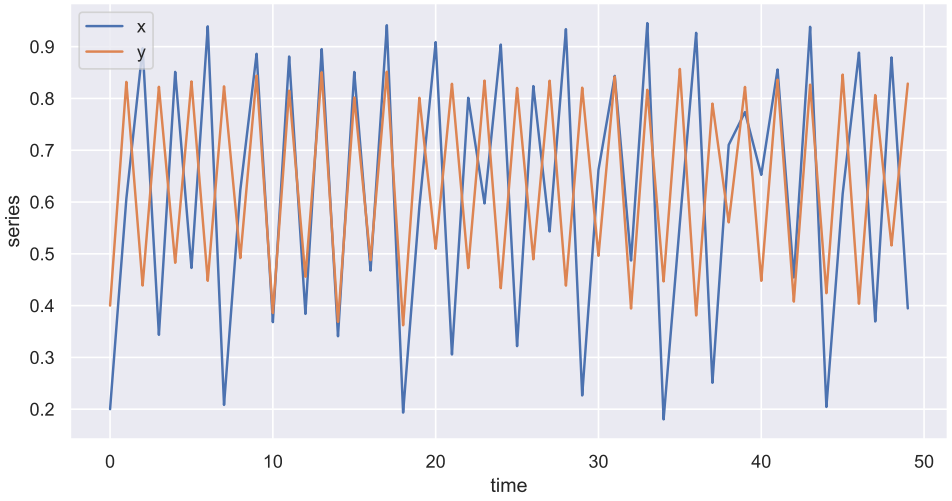

Figure 2.1: Coupled logistic time series.

2.2.2 Lorenz

One of the most intensively studied dynamical systems associated with chaotic behavior is the Lorenz attractor. Due to its shape, it has been widely associated with the famous butterfly effect known from chaos theory. The system itself is defined by three ODEs given as:

$$\begin{aligned} \frac{dx}{dt} &= \sigma \cdot (y - x) \\ \frac{dy}{dt} &= x \cdot (\rho - z) - y \\ \frac{dz}{dt} &= x \cdot y - \beta \cdot z \,. \end{aligned} \tag{2.3}$$

The parameters originally studied by Lorenz are set at $\sigma = 10$, $\beta = 8/3$, and $\rho = 28$, wherefore the system is in the chaotic regime and explicitly contains nonlinear correlations. If not stated otherwise, we simulate the equations in timesteps $dt = 0.01$ starting from the initial condition $(x_1, y_1, z_1) = (1, 1, 1)$. This parametrization ensures a smooth resolution of the attractor as depicted in Figure 2.2.

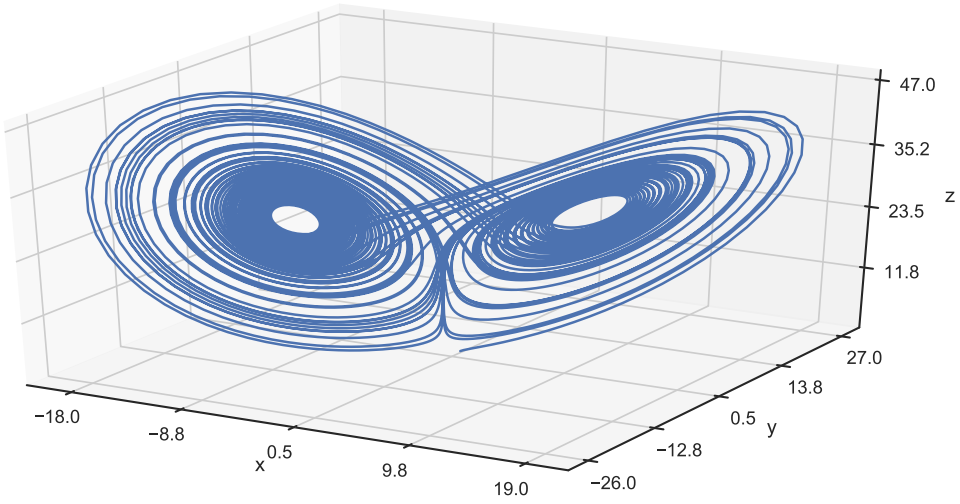

Figure 2.2: Lorenz attractor.

2.2.3 Rössler

While the Lorenz attractor builds on simplified physics models including plasmas and dynamos the Rössler attractor was solely developed for academic purposes. It explicitly only exhibits nonlinearity between its x and z coordinates encoded in the following ODEs:

$$\begin{aligned} \frac{dx}{dt} &= -y - z \\ \frac{dy}{dt} &= x + a \cdot y \\ \frac{dz}{dt} &= b + x \cdot (z - c)\,, \end{aligned} \tag{2.4}$$

where chaotic behavior is revealed for $a = 0.2$, $b = 0.2$, and $c = 5.7$. We simulate the equations in timesteps $dt = 0.01$ starting from the initial condition $(x_1, y_1, z_1) = (2, 2, 2)$. The attractor for this configuration is depicted in Figure 2.3.

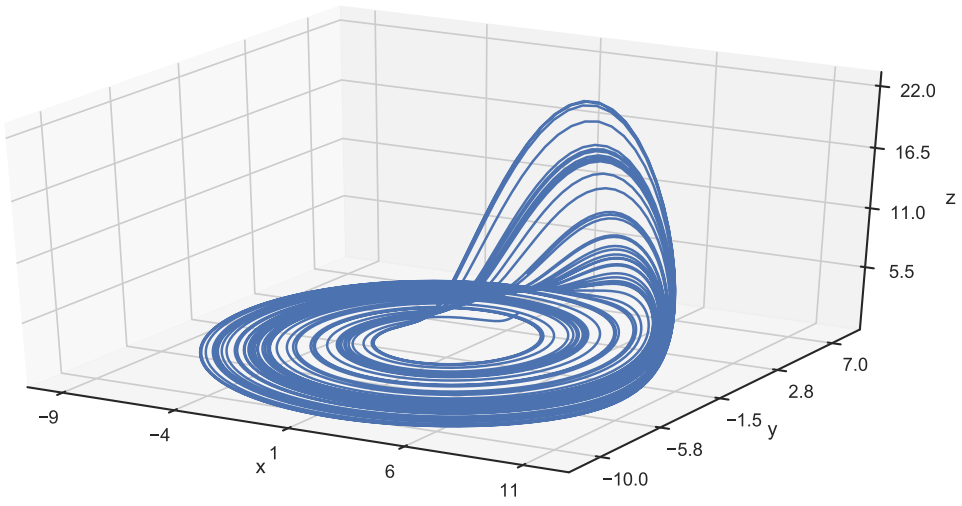

Figure 2.3: Rössler attractor.

2.2.4 Dummy Systems

In order to control for causality arising from singularities in our model, we build synthetic dummy systems which mimic periodic and random behavior.

Coupled Sinus

The first control system we introduce is a coupled sinus system. It serves as a dummy for seasonal and lagged time series systems:

$$\begin{aligned} x_t &= \sin(t) \\ y_t &= r_{xy} \cdot x_{t-\tau_y} = r_{xy} \cdot \sin(t - \tau_y) \\ z_t &= y_t - r_{xz} \cdot x_{t-\tau_z} = y_t - r_{xz} \cdot \sin(t - \tau_z)\,, \end{aligned} \tag{2.5}$$

where we choose $\tau_y = 2$, $\tau_z = 2.5$ $r_{xy} = 1.2$, and $r_{xz} = 0.4$ as the default setting.

Random Gaussian

In order to rule out spurious statistical effects, we create a coupled random model:

$$\begin{aligned} x_t &= W_x(t) \\ y_t &= W_y(t) \\ z_t &= r_x \cdot W_x(t) + r_y \cdot W_y(t)\,, \end{aligned} \tag{2.6}$$

where $W_x(t)$ and $W_y(t)$ are Wiener processes and $r_x = r_y = 0.5$ is our default parametrization.

2.3 Preprocessing

Even though we ensure cleanliness of our data in the collection process, further steps are needed before we begin running analyses on our time series systems. We preprocess the raw data using similar techniques and parameters as Haluszczynski et al. [2].

2.3.1 Fixed Windows

In order to obtain dynamically evolving results, we divide the time series $x_{n,\underline{T}}$ in overlapping windows. To do so, we select a fixed-size sliding window of $T_w < T$ time steps. The step size between two consecutive windows is set at δT. Accordingly, the first window is given by $x_{n,\underline{T},1} = (x_{n,1}, \ldots, x_{n,T_w})$, the second window by $x_{n,\underline{T},2} = (x_{n,1+\delta T_w}, \ldots, x_{n,T_w+\delta T})$, and so forth. In general, the w-th window is:

$$x_{n,\underline{T},w} \equiv x_{n,1+(w-1)\cdot\delta T:T_w+(w-1)\cdot\delta T} = \left(x_{n,1+(w-1)\cdot\delta T}, \ldots, x_{n,T_w+(w-1)\cdot\delta T}\right).$$

The total number of windows W is given by:

$$W = \left\lfloor \frac{T - T_w}{\delta T} \right\rfloor + 1 .$$

For an N-dimensional system the w-th window is:

$$x_{\underline{N},\underline{T},w} \equiv \left(x_{n,T,w} \mid n \in \{1, \ldots, N\}\right),$$

or written as a matrix:

$$x_{\underline{N},\underline{T},w} \equiv \begin{pmatrix} x_{1,1+(w-1)\cdot\delta T} & x_{1,2+(w-1)\cdot\delta T} & \cdots & x_{1,T_w+(w-1)\cdot\delta T} \\ x_{2,1+(w-1)\cdot\delta T} & x_{2,2+(w-1)\cdot\delta T} & \cdots & x_{2,T_w+(w-1)\cdot\delta T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1+(w-1)\cdot\delta T} & x_{N,2+(w-1)\cdot\delta T} & \cdots & x_{N,T_w+(w-1)\cdot\delta T} \end{pmatrix} \in \mathbb{R}^{N\times T_w} . \tag{2.7}$$

2.3.2 Flexible Windows

While we primarily focus on analyzing the evolution of causality measures, we furthermore investigate whether the results are robust for different time series lengths. Therefore, we divide the time series $x_{n,\underline{T}}$ into flexible-sized windows, which are all aligned at $t = 1$. We preset a minimum and maximum window size $T_{min} < T_{max} \leq T$ and a number of windows W. Thus, the step between two lengths is:

$$\delta T = \left\lfloor \frac{T_{max} - T_{min}}{W - 1} \right\rfloor , \tag{2.8}$$

and hence, the first window is given by $x_{n,\underline{T},1} = (x_{n,1}, \ldots, x_{n,\delta T})$. In general, the w-th window of a single time series is:

$$x_{n,\underline{T},w} \equiv x_{n,1:T_{min}+(w-1)\cdot\delta T} = \left(x_{n,1}, \ldots, x_{n,T_{min}+(w-1)\cdot\delta T}\right).$$

In order to align our notation with the fixed-sized windows, we denote the size of the w-th window analogously as $T_w \equiv T_{min} + (w-1)\cdot\delta T$. For an N-dimensional system the w-th flexible window in matrix form is:

$$x_{\underline{N},\underline{T},w} \equiv \begin{pmatrix} x_{1,1:T_w} \\ x_{2,1:T_w} \\ \vdots \\ x_{N,1:T_w} \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,T_w} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,T_w} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,T_w} \end{pmatrix} \in \mathbb{R}^{N\times T_w} . \tag{2.9}$$

2.3.3 Rank-Ordered Remapping

In order to compensate for effects from static nonlinearities, we perform a Gaussian rank-ordered remapping of the time series. For a given time series $x_{n,\underline{T}}$ of length T, the procedure requires four steps:

1. Firstly, we draw a corresponding series of T Gaussian distributed random numbers $g_{n,\underline{T}}$. To do so, we take samples from the standard normal distribution $\mathcal{N}(0,1)$.

2. We rearrange $g_{n,\underline{T}}$ in ascending order, so that $g^{\uparrow}_{n,s} \leq g^{\uparrow}_{n,t}$ for $s \leq t$ and analogously reorder $x_{n,\underline{T}}$, to $x^{\uparrow}_{n,\underline{T}}$.

3. Furthermore, we define a ranking map φ, which delivers a sequence of the ranking[1] of every element in $x_{n,\underline{T}}$ by mapping it to its ordered version $x^{\uparrow}_{n,\underline{T}}$:

$$\begin{aligned} \varphi : \mathbb{R}^T \times \mathbb{R}^T &\longrightarrow \mathrm{Sym}(\{1,\dots,T\}) \\ \left(x_{n,\underline{T}}, x^{\uparrow}_{n,\underline{T}}\right) &\longmapsto r_{n,\underline{T}}\,, \end{aligned} \tag{2.10}$$

   where $\mathrm{Sym}(M)$ is the symmetry or permutation group of a given set M. For example, if the t-th element of $x_{n,\underline{T}}$ was the i-th lowest in the time series, then $r_{n,t} = i$, and so forth.

4. The final step is to rearrange the ordered Gaussian series according to the ranking map φ. Hence, the remapped time series[2] is:

$$g^{\uparrow}_{n,\varphi\left(x_{n,\underline{T}},x^{\uparrow}_{n,\underline{T}}\right)}\,. \tag{2.11}$$

Simply put, the rank-ordered remapping algorithm replaces the lowest element of a time series with the lowest value of the corresponding Gaussian series and so forth. This allows us to trace back conclusions purely to dynamic nonlinearities since the distributions are Gaussian for both the time series and its surrogate. We thoroughly discuss time series surrogates in Section 2.4.

[1] Note that at this point we assume that all elements in $x_{n,\underline{T}}$ and $g_{n,\underline{T}}$ are unique. If this is not the case, we can simply add infinitesimal small numbers to make the elements distinct.

[2] Due to our notation established in Section 2.1, this expression directly translates to a surrogate time series system by capitalizing $n \to N$.

2.3.4 Rescaling

As an alternative to rank-ordered remapping, we rescale the time series to a fixed interval $[s_{min}, s_{max}]$ which conserves the shape of its distribution. Subsequently, for a time series $x_{n,\underline{T}}$ with $x_{n,min}$ and $x_{n,max}$ being its minimum and maximum values, every rescaled element of the rescaled series is given by:

$$\begin{aligned} &\phi : \mathbb{R} \longrightarrow [s_{min}, s_{max}] \\ &(x_{n,t}) \longmapsto s_{n,t} \equiv \frac{s_{max} - s_{min}}{x_{n,max} - x_{n,min}} \cdot (x_{n,t} - x_{n,min}) + s_{min} \,. \end{aligned} \tag{2.12}$$

We directly obtain the expression for the whole rescaled time series system by capitalizing both indices to $s_{\underline{N},\underline{T}}$. Our default scaling interval is $[s_{min}, s_{max}] = [0, 1]$.

2.4 Surrogates

One of the central questions of this thesis is whether we can find a connection between nonlinearity and causality. Therefore, we need to destroy the nonlinear properties of the time series, which is achieved by *Fourier Transform* (FT) surrogates. We discuss two of the most common techniques and refer to Räth and Monetti [3] for further details.

2.4.1 Fourier Transform

The most basic form of surrogatization for a time series $x_{n,\underline{T}}$ is performed according to the following three steps:

1. We begin by Fourier transforming the time series:

   $$x_{n,\underline{\Omega}} = \mathcal{F}\{x_{n,\underline{T}}\} \,,$$

   which separates all linear properties into the amplitudes while keeping the nonlinear properties in the phases.

2. In order to diminish the nonlinear properties we add a set of uniformly distributed random numbers to the Fourier phases:

   $$x_{n,\underline{\Omega}} \cdot e^{i\phi_k} \,,$$

   where $k \in \{1, \ldots, K\}$ denotes k-th set of random numbers and $\phi_k \in [0, 2\pi]$.

3. Finally, we perform an inverse Fourier transformation and obtain our k-th surrogate realization:
$$\tilde{x}^{(k)}_{n,\underline{T}} = \mathcal{F}^{-1}\left\{x_{n,\underline{\Omega}} \cdot e^{i\phi_k}\right\}.$$

We apply the same set of random phases ϕ_k to all time series in a system $x_{\underline{N},\underline{T}}$ and hence we denote $\tilde{x}^{(k)}_{\underline{N},\underline{T}}$ as the FT surrogate system. In order to ensure stability of our results, we generally use $K \geq 20$ realizations and average over the calculated statistical measures.

2.4.2 Amplitude Adjusted Fourier Transform

The drawback of ordinary FT surrogates is that the time series (amplitude) distributions are not preserved. This is addressed by *Amplitude Adjusted Fourier Transform* (AAFT) through inserting a Gaussian remapping step before and after the FT surrogatization. Hence, the procedure for a time series $x_{n,\underline{T}}$ is:

1. We perform a rank-ordered remapping as discussed in Subsection 2.3.3:
$$g^{\uparrow}_{n,\varphi\left(x_{n,\underline{T}},x^{\uparrow}_{n,\underline{T}}\right)}.$$

2. Then, we calculate the FT surrogate of the remapped time series $g^{\uparrow}_{n,\varphi\left(x_{n,\underline{T}},x^{\uparrow}_{n,\underline{T}}\right)}$ as explained in Subsection 2.4.1:
$$\tilde{g}_{n,\underline{T}} = \mathcal{F}^{-1}\left\{g^{\uparrow}_{n,\varphi\left(x_{n,\underline{\Omega}},x^{\uparrow}_{n,\underline{\Omega}}\right)} \cdot e^{i\phi_k}\right\},$$
where we omit the arrow in the superscript as the series is not ordered anymore.

3. Finally, we map the surrogate back to the original time series:
$$\tilde{x}^{(k)}_{n,\underline{T}} = x_{n,\varphi\left(\tilde{g}_{n,\underline{T}},\tilde{g}^{\uparrow}_{n,\underline{T}}\right)}.$$

While the distribution of the time series is conserved, this generally does not apply to its power spectrum. The tradeoff between conserving either is addressed by iteratively matching the power spectrum and the time series distribution of the surrogate time series to its original. We refer to Schreiber and Schmitz [4] for a thorough discussion on this procedure, which is called *Iterative Amplitude Adjusted Fourier Transform* (IAAFT).

Chapter 3

Time Series Measures

Before introducing inference methods of causality, we give a brief overview of statistical time series measures. To do so, we firstly introduce general definitions before we describe the relevant algorithms for this thesis.

3.1 General Concept

We begin by providing fundamental definitions for time series measures. This comprises setting up a general concept and the introduction of several measure classes.

3.1.1 Measure Systems

Before we discuss the respective calculation algorithms, we must establish a general concept of time series measures. For two time series $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$, let ψ be a bivariate measure:

$$\begin{aligned} \psi : \mathbb{R}^T \times \mathbb{R}^T &\longrightarrow \mathbb{R} \\ (x_{m,\underline{T}}, x_{n,\underline{T}}) &\longmapsto \psi \,. \end{aligned} \tag{3.1}$$

As illustrated in Section 2.1 we define the corresponding measure system $\psi_{\underline{N},\underline{N}}$ for an N-dimensional system $x_{N\times T}$ with length T:

$$\begin{aligned} \psi_{\underline{N},\underline{N}} : \mathbb{R}^{N\times T} &\longrightarrow \mathbb{R}^{N\times N} \\ (x_{\underline{N},\underline{T}}) &\longmapsto \psi_{\underline{N},\underline{N}} \equiv \left\{ \psi(x_{m,\underline{T}}, x_{n,\underline{T}}) \mid m,n \in \{1,\ldots,N\} \right\}, \end{aligned} \tag{3.2}$$

or written as a matrix:

$$\psi_{\underline{N},\underline{N}} \equiv \begin{pmatrix} \psi(x_{1,\underline{T}}, x_{1,\underline{T}}) & \psi(x_{1,\underline{T}}, x_{2,\underline{T}}) & \dots & \psi(x_{1,\underline{T}}, x_{N,\underline{T}}) \\ \psi(x_{2,\underline{T}}, x_{1,\underline{T}}) & \psi(x_{2,\underline{T}}, x_{2,\underline{T}}) & \dots & \psi(x_{2,\underline{T}}, x_{N,\underline{T}}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi(x_{N,\underline{T}}, x_{1,\underline{T}}) & \psi(x_{N,\underline{T}}, x_{2,\underline{T}}) & \dots & \psi(x_{N,\underline{T}}, x_{N,\underline{T}}) \end{pmatrix} . \tag{3.3}$$

We extend the notation for time series to measure systems by using corresponding indices:

- The element in the m-th row and n-th column of the matrix is denoted by:

$$\psi_{m,n} = \psi(x_{m,\underline{T}}, x_{n,\underline{T}}) \in \mathbb{R} .$$

- The m-th row of the matrix is denoted by:

$$\psi_{m,\underline{N}} \equiv \big(\psi(x_{m,\underline{T}}, x_{1,\underline{T}}), \psi(x_{m,\underline{T}}, x_{2,\underline{T}}), \dots, \psi(x_{m,\underline{T}}, x_{N,\underline{T}})\big) \in \mathbb{R}^N .$$

- The n-th column of the matrix is denoted by:

$$\psi_{\underline{N},n} \equiv \begin{pmatrix} \psi(x_{1,\underline{T}}, x_{n,\underline{T}}) \\ \psi(x_{2,\underline{T}}, x_{n,\underline{T}}) \\ \vdots \\ \psi(x_{N,\underline{T}}, x_{n,\underline{T}}) \end{pmatrix} \in \mathbb{R}^N .$$

Thus, we can simply extend an arbitrary bivariate statistical measure ψ to a measure system matrix by adding two underlined indices to its symbol $\psi_{\underline{N},\underline{N}}$.

3.1.2 Composite Measures

For two time series $x_{n,\underline{T}}$, $x_{m,\underline{T}}$, let ψ^i and ψ^j be two bivariate measures. Then we can compose a third measure:

$$\begin{aligned} \psi : \mathbb{R}^T \times \mathbb{R}^T &\longrightarrow \mathbb{R} \\ (x_{m,\underline{T}}, x_{n,\underline{T}}) &\longmapsto \psi \equiv \psi^i(x_{m,\underline{T}}, x_{n,\underline{T}}) \otimes \psi^j(x_{m,\underline{T}}, x_{n,\underline{T}}) , \end{aligned} \tag{3.4}$$

where $\otimes$ denotes an ordinary operator on $\mathbb{R}$, such as multiplication or addition. This principle can simply be extended to any number of measures using iteration. We can introduce an arbitrary third bivariate measure ψ^k, compose a new measure $\psi^l = \psi \otimes \psi^k = \psi^i \otimes \psi^j \otimes \psi^k$, and so forth.

3.1.3 Surrogate Measures

Hitherto, we have only discussed measures calculated on time series and not on their surrogates. We recall Section 2.4, where we expressed the k-th surrogate realization of a time series $x_{n,\underline{T}}$ as $\tilde{x}^{(k)}_{n,\underline{T}}$. Hence, we denote a statistical measure ψ between two surrogate time series $\tilde{x}^{(k)}_{m,\underline{T}}$ and $\tilde{x}^{(k)}_{n,\underline{T}}$ as follows:

$$\tilde{\psi}^{(k)}(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \psi\left(\tilde{x}^{(k)}_{m,\underline{T}}, \tilde{x}^{(k)}_{n,\underline{T}}\right). \tag{3.5}$$

As indicated, we ensure robustness of our results by averaging over the surrogate measures for multiple realizations $k \in \{1, \ldots, K\}$. For a bivariate surrogate measure we denote the mean as:

$$\tilde{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \left\langle \tilde{\psi}^{(k)}(x_{m,\underline{T}}, x_{n,\underline{T}}) \right\rangle_k = \frac{1}{K}\sum_{k=1}^{K} \tilde{\psi}^{(k)}(x_{m,\underline{T}}, x_{n,\underline{T}}), \tag{3.6}$$

with its corresponding system matrix $\tilde{\psi}_{\underline{N},\underline{N}}$.

The associated sample standard deviation of the surrogate measure is:

$$\begin{aligned} \tilde{\psi}^{\sigma}(x_{m,\underline{T}}, x_{n,\underline{T}}) &\equiv \sqrt{\left\langle \left(\tilde{\psi}^{(k)}(x_{m,\underline{T}}, x_{n,\underline{T}}) - \tilde{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}})\right)^2 \right\rangle_k} \\ &= \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\tilde{\psi}^{(k)}(x_{m,\underline{T}}, x_{n,\underline{T}}) - \tilde{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}})\right)^2}, \end{aligned} \tag{3.7}$$

with its corresponding system matrix $\tilde{\psi}^{\sigma}_{\underline{N},\underline{N}}$.

3.1.4 Nonlinear Measures

As discussed in Section 2.4, surrogate time series destroy the nonlinear properties of time series while keeping the linearities intact. Following this logic, a measure ψ incorporates contributions from linear properties, which is $\tilde{\psi}$, and an unknown nonlinear rest.

Hence, we calculate the nonlinear part denoted as:

$$\Delta\psi(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \psi(x_{m,\underline{T}}, x_{n,\underline{T}}) - \tilde{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}}), \tag{3.8}$$

which we refer to as the difference nonlinearity of ψ.

Since we rule out negative nonlinearities attributing them to spurious effects, we take the absolute value or, as an alternative, apply the maximum function to the difference nonlinearity. In extension to Haluszczynski et al. [2], we introduce the following nonlinearity measures which quantify the fraction of nonlinearity within arbitrary ψ:

- The relative nonlinearity is defined as:

$$\dot{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \frac{\Delta\psi(x_{m,\underline{T}}, x_{n,\underline{T}})}{\psi(x_{m,\underline{T}}, x_{n,\underline{T}})} \,. \tag{3.9}$$

- As we rule out negative nonlinearities attributing them to spurious effects, we specify the absolute relative nonlinearity as:

$$\dot{\psi}^{+}(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \frac{|\Delta\psi(x_{m,\underline{T}}, x_{n,\underline{T}})|}{\psi(x_{m,\underline{T}}, x_{n,\underline{T}})} \,. \tag{3.10}$$

- The maximum nonlinearity is given by:

$$\dot{\psi}^{max}(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \frac{\max\{0, \Delta\psi(x_{m,\underline{T}}, x_{n,\underline{T}})\}}{\psi(x_{m,\underline{T}}, x_{n,\underline{T}})} \,, \tag{3.11}$$

  which is normalized to $[0, 1]$ for positive definite measures ψ.

3.1.5 Cross Measures

In order to directly capture the nonlinear contribution of a time series to another, we introduce cross measures. In case of directional measures, they quantify how much the linear part of a time series $x_{m,\underline{T}}$ influences another series $x_{n,\underline{T}}$ under the measure ψ:

$$\check{\psi}^{(k)}(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \psi\left(\tilde{x}^{(k)}_{m,\underline{T}}, x_{n,\underline{T}}\right) . \tag{3.12}$$

In particular, we are interested in cross measures of a time series $x_{n,\underline{T}}$ with itself:

$$\check{\psi}^{(k)}(x_{n,\underline{T}}, x_{n,\underline{T}}) \equiv \psi\left(\tilde{x}^{(k)}_{n,\underline{T}}, x_{n,\underline{T}}\right) . \tag{3.13}$$

This quantifies how much information within $x_{n,\underline{T}}$ is encoded in its linear part under the measure ψ and serves as an indicator of how much linearity is present in a time series.

Analogous to the surrogate measures, we denote the mean as $\check{\psi}$ and the standard deviation as $\check{\psi}^{\sigma}$. Hence, we introduce the difference cross nonlinearity:

$$\Delta\check{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \psi(x_{m,\underline{T}}, x_{n,\underline{T}}) - \check{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}})\,, \tag{3.14}$$

Furthermore, we define the surrogate difference cross nonlinearity:

$$\tilde{\Delta}\check{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \check{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}}) - \tilde{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}})\,, \tag{3.15}$$

which quantifies how much of the linearity in $x_{n,\underline{T}}$ is attributed to the linear part in $x_{m,\underline{T}}$. Note that in contrast to the ordinary difference cross nonlinearity, the terms are switched.

Based on these two difference measures, we calculate the nonlinearity fractions defined in Subsection 3.1.4. While we divide by $\psi(x_{m,\underline{T}}, x_{n,\underline{T}})$ for the ordinary cross measure, we use $\check{\psi}(x_{m,\underline{T}}, x_{n,\underline{T}})$ for its corresponding surrogate.

3.1.6 Measure Evolutions

In Subsections 2.3.1 and 2.3.2 we described how to divide a time series into fixed- or flexible-size windows. Thus, instead of calculating measures on the whole time series, we compute them on the individual windows.

Let ψ be the bivariate measure for two time series $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$, then the measure calculated on the w-th window is:

$$\begin{aligned} \psi_w : \mathbb{R}^{T_w} \times \mathbb{R}^{T_w} &\longrightarrow \mathbb{R} \\ (x_{m,\underline{T},w}, x_{n,\underline{T},w}) &\longmapsto \psi_w \equiv \psi(x_{m,\underline{T},w}, x_{n,\underline{T},w})\,, \end{aligned} \tag{3.16}$$

with its corresponding measure system:

$$\begin{aligned} \psi_{\underline{N},\underline{N},w} : \mathbb{R}^{N\times T_w} &\longrightarrow \mathbb{R}^{N\times N} \\ (x_{\underline{N},\underline{T},w}) &\longmapsto \psi_{\underline{N},\underline{N},w} \equiv \left\{\psi_w(x_{m,\underline{T},w}, x_{n,\underline{T},w}) \mid m,n \in \{1,\ldots,N\}\right\}. \end{aligned} \tag{3.17}$$

We define a tensor in order to aggregate the measure systems for all windows $w \in \{1,\ldots,W\}$ into one expression:

$$\begin{aligned} \psi_{\underline{N},\underline{N},\underline{W}} : \mathbb{R}^{N\times(T_1+\ldots+T_w)} &\longrightarrow \mathbb{R}^{N\times N\times W} \\ (x_{\underline{N},\underline{T},1},\ldots,x_{\underline{N},\underline{T},W}) &\longmapsto \psi_{\underline{N},\underline{N},\underline{W}} \equiv \left\{\psi_{\underline{N},\underline{N},w} \mid w \in \{1,\ldots,W\}\right\}. \end{aligned} \tag{3.18}$$

We henceforth refer to vectors along the third w-dimension as evolutions. Assuming $\psi(x_{m,\underline{T}}, x_{n,\underline{T}})$ is an asymmetric measure quantifying the influence of X_m on X_n for better clarity, we make the following interpretation suggestions:

- The vector $\psi_{\underline{N},n,w}$ denotes the influence which all the state variables in the system have on X_n within the w-th window:

$$\psi_{\underline{N},n,w} \equiv \begin{pmatrix} \psi_{1,n,w} \\ \psi_{2,n,w} \\ \vdots \\ \psi_{N,n,w} \end{pmatrix} \in \mathbb{R}^N .$$

- The vector $\psi_{m,\underline{N},w}$ denotes the influence which X_m has on all the other variables in the system within the w-th window:

$$\psi_{m,\underline{N},w} \equiv (\psi_{m,1,w}, \psi_{m,2,w}, \ldots, \psi_{m,N,w}) \in \mathbb{R}^N .$$

- The vector $\psi_{m,n,W}$ denotes the evolution of the influence which X_m has on X_n:

$$\psi_{m,n,\underline{W}} \equiv (\psi_{m,n,1}, \psi_{m,n,2}, \ldots, \psi_{m,n,W}) \in \mathbb{R}^W .$$

  In cases where m and n are specified, we reduce the expression to $\psi_{\underline{W}}$.

- The matrix $\psi_{\underline{N},\underline{N},w}$ denotes the system influence within the w-th window:

$$\psi_{\underline{N},\underline{N},w} \equiv \begin{pmatrix} \psi_{1,1,w} & \psi_{1,2,w} & \cdots & \psi_{1,N,w} \\ \psi_{2,1,w} & \psi_{2,2,w} & \cdots & \psi_{2,N,w} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{N,1,w} & \psi_{N,2,w} & \cdots & \psi_{N,N,w} \end{pmatrix} \in \mathbb{R}^{N \times N} .$$

- The matrix $\psi_{\underline{N},n,\underline{W}}$ denotes the evolution of the influence which all the state variables in the system have on X_n:

$$\psi_{\underline{N},n,\underline{W}} \equiv \begin{pmatrix} \psi_{1,n,1} & \psi_{1,n,2} & \cdots & \psi_{1,n,W} \\ \psi_{2,n,1} & \psi_{2,n,2} & \cdots & \psi_{2,n,W} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{N,n,1} & \psi_{N,n,2} & \cdots & \psi_{N,n,W} \end{pmatrix} \in \mathbb{R}^{N \times W} .$$

- The matrix $\psi_{m,\underline{N},\underline{W}}$ denotes the evolution of the influence which X_m has on all the variables in the system:

$$\psi_{m,\underline{N},\underline{W}} \equiv \begin{pmatrix} \psi_{m,1,1} & \psi_{m,1,2} & \cdots & \psi_{m,1,W} \\ \psi_{m,2,1} & \psi_{m,2,2} & \cdots & \psi_{m,2,W} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{m,N,1} & \psi_{m,N,2} & \cdots & \psi_{m,N,W} \end{pmatrix} \in \mathbb{R}^{N \times W} .$$

3.1.7 Nested Evolution Measures

One of our research questions is to find relationships between statistical and causality measures. Therefore, we analyze relationships between different measure evolutions.

Let ψ be a bivariate measure between two time series as specified in equation 3.1. Furthermore, let $\psi^j_{m,n,\underline{W}}$ and $\psi^j_{m,n,\underline{W}}$ denote the evolutions of two different measures between the state variables X_m and X_n in a system. Then, we calculate:

$$\begin{aligned} \psi : \mathbb{R}^W \times \mathbb{R}^W &\longrightarrow \mathbb{R} \\ \left(\psi^i_{m,n,\underline{W}}, \psi^j_{m,n,\underline{W}}\right) &\longmapsto \psi \equiv \psi\left(\psi^i_{m,n,\underline{W}}, \psi^j_{m,n,\underline{W}}\right), \end{aligned} \tag{3.19}$$

and analogously its surrogate $\tilde{\psi}$. We henceforth call this a nested evolution measure.

3.2 Measures Algorithms

After setting up general definitions of time series measures, we in the following discuss the algorithms of relevant measures for this thesis. We distinguish univariate, bivariate, and system measures according to the number of time series they are calculated on.

3.2.1 Univariate Measures

As discussed, univariate measures can be calculated on one single time series. While we defined the general definitions in Section 3.1 for bivariate time series, they are directly applicable to the univariate case.

Mean

The mean of a time series $x_{n,\underline{T}}$ is defined as:

$$\mu(x_{n,\underline{T}}) \equiv \frac{1}{T} \sum_{t=1}^{T} x_{n,t}\,, \tag{3.20}$$

with its corresponding mean matrix denoted as $\mu_{\underline{N}}$. The mean evolution is expressed as $\mu_{\underline{N},\underline{W}}$.

Standard Deviation

The sample variance of a time series $x_{n,\underline{T}}$ is defined as:

$$\sigma(x_{n,\underline{T}}) \equiv \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(x_{n,t} - \mu(x_{n,\underline{T}}))^2}\,, \tag{3.21}$$

with its corresponding system and evolution measure denoted as $\sigma_{\underline{N}}$ and $\sigma_{\underline{N},\underline{W}}$, respectively.

Percentile Value

Given a percentile α, the corresponding percentile value of a time series $x_{n,\underline{T}}$ is defined as the $\lfloor \alpha \cdot T \rfloor$-th value in the ordered series $x^{\uparrow}_{n,\underline{T}}$:

$$\mu^{\alpha}(x_{n,\underline{T}}) = x^{\uparrow}_{n,\lfloor \alpha \cdot T \rfloor}\,. \tag{3.22}$$

For $\alpha = 0.5$ the percentile value is referred to as the median.

Probability Density Estimation

Before we introduce information-theoretic statistical measures, we hereby briefly discuss two approaches to estimate the *Probability Density Function* (PDF) of a time series:

- The computationally most efficient method to estimate a discrete PDF is the histogram, which creates a discrete distribution by sorting the time series values into bins. We will conduct our analyses for different choices for the number of bins n_b and bin ranges r_b.
- *Kernel Density Estimation* (KDE) is a nonparametric technique to estimate the continuous PDF by overlapping positive and symmetric kernel functions K:

$$\hat{p}(x_{n,\underline{T}}) = \frac{1}{T \cdot b}\sum_{t=1}^{T} K\left(\frac{x - x_{n,t}}{b}\right), \tag{3.23}$$

  where b is a smoothing parameter called bandwidth. As a default setting we use the Gaussian kernel and estimate the bandwidth using Scott's rule:

$$b = T^{-\frac{1}{d+4}}\,,$$

where d is the dimension of the distribution to be approximated.

Both techniques can be easily extended to multiple time series or dimensions, which will become relevant in the upcoming section.

Entropy

The marginal entropy H of a time series $x_{n,\underline{T}}$ can be calculated through:

$$H(x_{n,\underline{T}}) = -\sum_{t=1}^{T} p(x_{n,t}) \cdot \log p(x_{n,t}), \tag{3.24}$$

where $p(x_{n,\underline{T}})$ denotes the discrete PDF.

In the continuous case, the summation over the time series elements is replaced by an integral. Henceforth, we will only give the definitions for discrete PDFs.

Stationarity

A time series is stationary if its PDF is constant over time. For a time series divided into W windows denoted as $x_{n,\underline{T},w}$, the condition of stationarity is given by:

$$p(x_{n,\underline{T},w}) = const \quad \forall w \in \{1, \ldots, W\}\,. \tag{3.25}$$

Convergence

We check the convergence of a time series by calculating the standard deviation of a time series for rolling windows similar to the flexible windows defined in Subsection 2.3.2. However, we fix the terminal value of the windows instead of the starting point. For a total number of W windows and $\delta T = \lfloor T/W \rfloor$, the w-th window is given by:

$$x_{n,\underline{T},w} \equiv x_{n,(w-1)\cdot\delta T:T} = \left(x_{n,(w-1)\cdot\delta T}, x_{n,(w-1)\cdot\delta T+1}, \ldots, x_{n,T}\right). \tag{3.26}$$

Subsequently we calculate the standard deviation for each window until it falls below a threshold θ:

$$\sigma(x_{n,\underline{T},w}) < \theta\,, \tag{3.27}$$

with a default value $\theta = 0.05$.

For large time series with $T > 100$, we use fixed-size windows as defined in Subsection 2.3.1 instead of flexible windows.

3.2.2 Bivariate Measures

The main type of statistical measures for this thesis are bivariate measures calculated between two time series. It is important to note that bivariate measures are not generally symmetric.

Pearson Correlation

Given two time series $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$, the *Pearson correlation coefficient* is calculated via:

$$\rho(x_{m,\underline{T}}, x_{n,\underline{T}}) = \frac{\sum_{t=1}^{T}(x_{m,t} - \mu(x_{m,\underline{T}})) \cdot (x_{n,t} - \mu(x_{n,\underline{T}}))}{\sqrt{\sum_{t=1}^{T}(x_{m,t} - \mu(x_{m,\underline{T}}))^2} \cdot \sqrt{\sum_{t=1}^{T}(x_{n,t} - \mu(x_{n,\underline{T}}))^2}}, \tag{3.28}$$

where $\bar{x}_{m,\underline{T}}$ and $\bar{x}_{n,\underline{T}}$ denote the respective time series means. It quantifies the direction and strength of their linear relationship by assigning a value within the interval $[-1, 1]$.

Correlation matrices $\rho_{\underline{N},\underline{N}}$ are symmetric and have diagonal values of 1.

Furthermore, as shown by Prichard and Theiler [5] they are equal to their corresponding surrogate correlation matrices $\tilde{\rho}_{\underline{N},\underline{N}}$, since the surrogatization only diminishes the nonlinear properties while leaving the linear correlation unaffected.

Mutual Information

Since correlations between time series are not merely linear, we utilize *Mutual Information* (MI) as a measure for both linear and nonlinear dependencies.

Analogously to the entropy defined in Subsection 3.2.1, the joint entropy between two time series $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$ is given by:

$$H(x_{m,\underline{T}}, x_{n,\underline{T}}) = -\sum_{t=1}^{T}\sum_{t=1}^{T} p(x_{m,t}, x_{n,t}) \cdot \log p(x_{m,t}, x_{n,t}), \tag{3.29}$$

where $p(x_{m,t}, x_{n,t})$ denotes the joint PDF.

The normalized mutual information for discrete PDFs is:

$$I(x_{m,\underline{T}}, x_{n,\underline{T}}) = \frac{H(x_{m,\underline{T}}) + H(x_{n,\underline{T}}) - H(x_{m,\underline{T}}, x_{n,\underline{T}})}{\sqrt{H(x_{m,\underline{T}}) \cdot H(x_{n,\underline{T}})}} \quad \in [0, 1]\,, \tag{3.30}$$

where 0 indicates no share of information between the series and 1 indicates identical PDFs.

Thus, normalized mutual information matrices $I_{\underline{N},\underline{N}}$ are symmetric and have diagonal values of 1.

A valuable property we utilize to validate our PDF-dependent measures is the similarity between the evolution of the Pearson correlation and the scaled surrogate mutual information. We check the similarity using the nested correlation:

$$\rho\Big(\rho_{\underline{W}}(x_{m,\underline{T}}, x_{n,\underline{T}}), \operatorname{sgn}_{\underline{W}}(\rho_{\underline{W}}(x_{m,\underline{T}}, x_{n,\underline{T}})) \odot \tilde{I}_{\underline{W}}(x_{m,\underline{T}}, x_{n,\underline{T}})\Big) \xrightarrow{W\to\infty} 1\,, \tag{3.31}$$

where $\operatorname{sgn}(*)$ is the signum function and $\odot$ denotes the Hadamard product.

Alternatively, we can check the equivalence between the arguments by additionally scaling the surrogate MI according to Subsection 2.3.4.

Nonlinear Correlation

Due to the similarity between surrogate MI and Pearson correlation, Haluszczynski et al. [2] introduced a measure capturing the nonlinear correlation calculated via:

$$\frac{\Big|I(x_{m,\underline{T}}, x_{n,\underline{T}}) - \tilde{I}(x_{m,\underline{T}}, x_{n,\underline{T}})\Big|}{I(x_{m,\underline{T}}, x_{n,\underline{T}})}\,, \tag{3.32}$$

which captures the proportion of mutual information driven by nonlinearities. Note that within our framework, nonlinear correlation is the absolute relative nonlinearity of the MI denoted as $\dot{I}^{+}(x_{m,\underline{T}}, x_{n,\underline{T}})$.

The corresponding nonlinear correlation matrices are asymmetric and have diagonal values of 0.

Conditional Entropy

The conditional entropy between two time series is given by:

$$H(x_{m,\underline{T}} \mid x_{n,\underline{T}}) = -\sum_{t=1}^{T}\sum_{t=1}^{T} p(x_{m,}, x_{n,t}) \cdot \log\frac{p(x_{m,t}, x_{n,t})}{p(x_{n,t})}\,, \tag{3.33}$$

which is generally an asymmetric measure.

3.2.3 System Measures

After having discussed how we obtain measure system matrices, we will now introduce methods on how to evaluate these matrices in one single value. Generally, we will dismiss the diagonal values of the measure matrices since they do not hold any relevant information about the system.

Mean

Given a measure system $\psi_{\underline{N},\underline{N}}$ with elements denoted as $\psi_{m,n}$, the measure system mean is calculated via:

$$\mu\big(\psi_{\underline{N},\underline{N}}\big) \equiv \frac{1}{N^2 - N} \sum_{m=1}^{N} \sum_{n=1}^{N} (1 - \delta_{m,n}) \cdot \psi_{m,n} \,, \tag{3.34}$$

where $\delta_{m,n}$ denotes the Kronecker delta.

In some cases we are interested in the mean of the absolute values, which is given by:

$$\mu^{+}\big(\psi_{\underline{N},\underline{N}}\big) \equiv \frac{1}{N^2 - N} \sum_{m=1}^{N} \sum_{n=1}^{N} (1 - \delta_{m,n}) \cdot |\psi_{m,n}| \,. \tag{3.35}$$

Geometric Mean

A related measure is the geometric mean of the system, which we compute via:

$$\mu^{g}\big(\psi_{\underline{N},\underline{N}}\big) \equiv \left(\prod_{m=1}^{N} \prod_{n=1}^{N} (1 - \delta_{m,n}) \cdot |\psi_{m,n}| + \delta_{m,n} \right)^{\frac{1}{N^2 - N}} \,. \tag{3.36}$$

Frobenius Norm

A common matrix measure is the Frobenius norm, which we slightly alter in order to exclude the diagonal values:

$$\big\|\psi_{\underline{N},\underline{N}}\big\|_F \equiv \sqrt{\sum_{m=1}^{N} \sum_{n=1}^{N} (1 - \delta_{m,n}) \cdot |\psi_{m,n}|^2} \,. \tag{3.37}$$

Asymmetry

We quantify the asymmetry of a measure matrix using several approaches:

- The difference asymmetry is defined as:

$$\Lambda\big(\psi_{\underline{N},\underline{N}}\big) \equiv \frac{1}{N^2 - N} \sum_{m=1}^{N} \sum_{n=1}^{N} (1 - \delta_{m,n}) \cdot (\psi_{m,n} - \psi_{m,n}) \,. \tag{3.38}$$

- In order to avoid cancellation between positive and negative values, the absolute difference asymmetry is defined as:

$$\Lambda^{+}\big(\psi_{\underline{N},\underline{N}}\big) \equiv \frac{1}{N^2 - N} \sum_{m=1}^{N} \sum_{n=1}^{N} (1 - \delta_{m,n}) \cdot |\psi_{m,n} - \psi_{m,n}| \,. \tag{3.39}$$

- We define the fractional asymmetry by comparing the symmetric and asymmetric part of a measure matrix, which is given by:

$$\frac{1}{2} \cdot \left(\psi_{\underline{N},\underline{N}} \pm \psi^{T}_{\underline{N},\underline{N}}\right), \tag{3.40}$$

  where the superscript T denotes the transposition of the matrix. Thus, the fractional asymmetry is defined as:

$$\Lambda^{f} \equiv \frac{\left\|\psi_{\underline{N},\underline{N}} - \psi^{T}_{\underline{N},\underline{N}}\right\|_{F}}{\left\|\psi_{\underline{N},\underline{N}} + \psi^{T}_{\underline{N},\underline{N}}\right\|_{F}} \,. \tag{3.41}$$

Chapter 4

Causality Measures

The concept of causality is deeply rooted in our way of thinking and is a fundamental principle of physics. In the following, we discuss several methods in order to infer causality within these systems.

4.1 General Concept

While some properties of causality, such as asymmetry, are common understanding among various disciplines, there exists no convention of requirements which a causality measure is supposed to fulfill. Hence, we hereby formulate constraints for the bivariate time series measure ψ from equation 3.1 in order to define a causality measure. Let $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$ be two distinct time series representing the m-th and n-th state variables in a time series system:

- *Asymmetry*: $\psi(x_{m,\underline{T}}, x_{n,\underline{T}})$ quantifies the direct causal influence of the m-th state variable X_m on the n-th variable X_n in the system. Hence, ψ is generally asymmetric: $\psi(x_{m,\underline{T}}, x_{n,\underline{T}}) \neq \psi(x_{n,\underline{T}}, x_{m,\underline{T}})$.
- *Normalization*: The codomain of ψ is bounded to a fixed interval:
  - In the case of non-directional causality measures, we normalize ψ so that $\psi(x_{m,\underline{T}}, x_{n,\underline{T}}) \in [0, 1]$. Hence, a value of 0 indicates no direct causal influence of $x_{m,\underline{T}}$ on $x_{n,\underline{T}}$, while a value 1 indicates strong causality.
  - For directional causality measures the codomain is set at $[-1, 1]$, where negative values indicate inverse causality in the same sense as negative correlation.

- *Integrity*: $\psi(x_{m,\underline{T}}, x_{m,\underline{T}})$ equals the same constant for any time series $x_{m,\underline{T}}$.

  In our case, the causality between two identical time series or state variables is either non existent for a value of 0 or maximally strong for 1. This highly depends on whether the inference method interprets causality in a time-lagged sense.

Furthermore, we make assumptions and interpretation suggestions for a causality system matrix $\psi_{\underline{N},\underline{N}}$:

- *Completeness*: The causality matrix $\psi_{\underline{N},\underline{N}}$ incorporates all direct bivariate causal relations, wherefrom any sub-relation can be deduced.
- *Activity*: The m-th row $\psi_{m,\underline{N}}$ denotes the direct causal influence, which X_m has on all the variables in the system.
- *Passivity*: The n-th column $\psi_{\underline{N},n}$ denotes the direct causal influence, which all state variables in the system have on X_n.

4.2 Granger Causality

One of the first practical approaches to measure causality was proposed by Clive Granger in 1969 [6]. *Granger Causality* (GC) quantifies causality following a temporal cause-effect intuition by analyzing whether a lagged version of a time series significantly influences another. The method relies on linear regression and thus only captures causality which stems from linear properties.

4.2.1 Ordinary Least Squares

For this reason we begin by giving a brief overview of *Ordinary Least Squares* (OLS), which is a technique for estimating parameters by minimizing the squared errors within a linear regression. It assumes that the response time series $x_{n,\underline{T}}$ is a linear function of the regressor series $x_{m,\underline{T}}$:

$$x_{n,t} = \sum_{t=1}^{T} \gamma_t \cdot x_{m,t} + \xi_t, \tag{4.1}$$

where γ_t is the n-dimensional vector of unknown parameters to be estimated and ξ_t denotes an independent error term.

Then, the OLS parameters can be obtained by minimizing the sum of squared errors such that errors of opposite signs are not cancelled out:

$$\hat{\gamma}_t = \arg\min_{\gamma_t} \sum_{t=1}^{T} (x_{n,t} - \gamma_t \cdot x_{m,t})^2 = \arg\min_{\gamma_t} \hat{\xi}_t^2. \tag{4.2}$$

We can quantify the accuracy of the OLS estimation using the *Residual Sum of Squares* (RSS):

$$\text{RSS} = \sum_{t=1}^{T} \hat{\xi}_t^2. \tag{4.3}$$

These fundamentals are essential for inferring GC as we will discuss in the following.

4.2.2 Regression Models

For two given time series $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$, X_m causes X_n if unique information in $x_{m,\underline{T}}$ exists, which is relevant for $x_{n,\underline{T}}$ and not contained in the past of $x_{n,\underline{T}}$.

Therefore, we firstly perform an auto-regression in order to find the relevant lag values of the time series $x_{n,\underline{T}}$:

$$\hat{x}_{n,t} = \sum_{\tau=1}^{\tau_{max}} \alpha_\tau \cdot x_{n,t-\tau} + \epsilon_t \tag{4.4}$$

where α_τ is the coefficient at lag τ, and ϵ_t denotes an independent error term. We call this regression the restricted model.

In the next step, we augment this regression using lag values of $x_{m,t}$

$$\hat{x}_{n,t} = \sum_{\tau=1}^{\tau_{max}} \alpha_\tau \cdot x_{n,t-\tau} + \sum_{\tau=1}^{\tau_{max}} \beta_\tau \cdot x_{m,t-\tau} + \eta_t, \tag{4.5}$$

where β_τ is the coefficient for the added time series at lag τ, and η_t denotes an independent error term. This is referred to as the augmented model.

Hence, Granger causality quantifies whether augmenting the auto-regression of $x_{n,\underline{T}}$ with past values from another time series $x_{m,\underline{T}}$ adds significant prediction value. If so, $x_{m,\underline{T}}$ is said to *Granger cause* $x_{n,\underline{T}}$. The significance of the causation is indicated using statistical hypothesis tests.

4.2.3 Statistical Hypothesis Test

The null hypothesis that $x_{m,\underline{T}}$ does not *Granger cause* $x_{n,\underline{T}}$ is formulated as:

$$H_0 : \beta_\tau = 0 \quad \forall\, \tau = \{1, \ldots, \tau_{max}\}.$$

We perform two different kinds of statistical tests in order to decide whether to accept or reject the null hypothesis.

F-Test

Most commonly, Granger causality is inferred using the F-Test, wherefore the test statistic is given by:

$$S_F = \frac{\mathrm{RSS}_{rest} - \mathrm{RSS}_{aug}}{\mathrm{RSS}_{aug}} \cdot \frac{T - 2 \cdot \tau_{max} - 1}{\tau_{max}}, \tag{4.6}$$

where RSS_{rest} and RSS_{aug} denote the RSS of the restricted and augmented model, respectively. The corresponding p-value is approximately calculated via:

$$p_F = 1 - F(S_F),$$

where $F(*)$ is the F-distribution.

Chi-Squared-Test

The statistic for the *Chi-Squared*-Test or χ^2-Test is computed similarly using:

$$S_{\chi^2} = \frac{\mathrm{RSS}_{rest} - \mathrm{RSS}_{aug}}{\mathrm{RSS}_{aug}} \cdot T. \tag{4.7}$$

Hence, the p-value under the χ^2-distribution is given by:

$$p_{\chi^2} = 1 - \chi^2\big(S_{\chi^2}\big).$$

4.2.4 Scoring

It is common practice to use the p-value of the statistical tests in order to assess the presence of GC between two time series. A p-value lower than a significance level α indicates that the null hypothesis defined in Subsection 4.2.3 can be rejected at the $(1 - \alpha)$-confidence level.

However, since we need an indicator for the strength of the causal coupling, we in the following propose a normalization scheme. In essence, GC measures the decrease in RSS when values of another time series are added to the autoregression. Therefore, we use the formula:

$$G(x_{m,\underline{T}}, x_{n,\underline{T}}) = 1 - \min\left\{ \left(\frac{\mathrm{RSS}_{aug}}{\mathrm{RSS}_{rest}} \right)^2, 1 \right\} \in [0,1]\,, \tag{4.8}$$

which will henceforth be our measure for GC.

We justify the validity of our formula using qualitative arguments:

- Given RSS_{rest}, if the error of the augmented model RSS_{aug} is smaller, then G approaches 1. Thus, G is 1 when the error of the augmented regression model is 0.
- If the error of the augmented model is larger, then $\min\left\{ \frac{\mathrm{RSS}_{aug}}{\mathrm{RSS}_{rest}}, 1 \right\} = 1$ and G is 0.
- If $x_{m,\underline{T}} = x_{n,\underline{T}}$, then $\mathrm{RSS}_{rest} = \mathrm{RSS}_{aug}$ and G is 0. We find this to be reasonable since no error reduction was achieved by augmenting the regression.
- We scale the fraction of the RSS quadratically, so that a smaller decrease in error leads to a higher increase in G.

Hence, the Granger causality G is normalized and fulfills the conditions of a causality measure as specified in Section 4.1.

Furthermore, since GC only captures causality from linear properties, GC matrices $G_{\underline{N},\underline{N}}$ are identical to their surrogate matrices $\tilde{G}_{\underline{N},\underline{N}}$.

4.3 Transfer Entropy

Transfer Entropy (TE) is an information-theoretical measure for quantifying the transfer of information from one time series to another. It was introduced by Schreiber [7] in order to address the drawback of MI, which includes shared information arising from common history and signals. Furthermore, TE can be interpreted as an extension of Granger causality and was shown by Barnett, Barrett, and Seth [8] to be equivalent to a factor of two for Gaussian random variables.

4.3.1 Conditional Mutual Information

As TE is mainly based on *Conditional Mutual Information* (CMI), we firstly provide the general expression of this statistical measure. Given three time series $x_{l,\underline{T}}$, $x_{m,\underline{T}}$, and $x_{n,\underline{T}}$, CMI is given by:

$$I(x_{l,\underline{T}}, x_{m,\underline{T}} \mid x_{n,\underline{T}}) = -\sum_{t=1}^{T}\sum_{t=1}^{T}\sum_{t=1}^{T} p(x_{m,t}, x_{n,t}, x_{l,t}) \cdot \log \frac{p(x_{n,t}) \cdot p(x_{l,t}, x_{m,t}, x_{n,t})}{p(x_{l,t}, x_{n,t}) \cdot p(x_{m,t}, x_{n,t})},$$

where $p(x_{m,t}, x_{n,t}, x_{l,t})$ is the joint PDF for three time series analogous to the bivariate case defined in Subsection 3.2.2.

Furthermore, we can express CMI in terms of entropy:

$$\begin{aligned} I(x_{l,\underline{T}}, x_{m,\underline{T}} \mid x_{n,\underline{T}}) = {} & H(x_{l,\underline{T}}, x_{n,\underline{T}}) + H(x_{m,\underline{T}}, x_{n,\underline{T}}) \\ & - H(x_{l,\underline{T}}, x_{m,\underline{T}}, x_{n,\underline{T}}) - H(x_{n,\underline{T}})\,. \end{aligned} \tag{4.9}$$

While MI can be normalized for discrete PDFs as mentioned in Subsection 3.2.2, this property has not been shown for CMI. Thus, we propose the following normalization:

$$I(x_{l,\underline{T}}, x_{m,\underline{T}} \mid x_{n,\underline{T}}) = \frac{H(x_{l,\underline{T}}, x_{n,\underline{T}}) + H(x_{m,\underline{T}}, x_{n,\underline{T}}) - H(x_{l,\underline{T}}, x_{m,\underline{T}}, x_{n,\underline{T}}) - H(x_{n,\underline{T}})}{\sqrt{H(x_{l,\underline{T}}, x_{n,\underline{T}}) \cdot H(x_{m,\underline{T}}, x_{n,\underline{T}})}}\,.$$

4.3.2 Scoring

Intuitively, TE quantifies the information gain past values of X_m deliver to future values of X_n without accounting for the history of X_n.

Thus, for a given lag $\tau \in \mathbb{N}^+$ TE is defined as:

$$T(x_{m,\underline{T}}, x_{n,\underline{T}}) = I(x_{n,\tau:T}, x_{m,1:T-\tau} \mid x_{n,1:T-\tau})\,. \tag{4.10}$$

Due to computational restrictions, we need to ensure that the time series arguments have equal length. However, while $x_{m,\underline{T}} \in \mathbb{R}^T$, $x_{m,1:T-\tau} \in \mathbb{R}^{T-\tau}$. Thus, by using $x_{m,\tau:T} \in \mathbb{R}^{T-\tau}$ we can express TE as:

$$\begin{aligned} T(x_{m,\underline{T}}, x_{n,\underline{T}}) &= I(x_{n,\tau:T}, x_{m,1:T-\tau} \mid x_{n,1:T-\tau}) \\ &= H(x_{n,\tau:T}, x_{n,1:T-\tau}) + H(x_{m,1:T-\tau}, x_{n,1:T-\tau}) \\ &\quad - H(x_{m,1:T-\tau}, x_{n,\tau:T}, x_{n,1:T-\tau}) - H(x_{n,1:T-\tau})\,, \end{aligned} \tag{4.11}$$

with its normalized form:

$$T(x_{m,\underline{T}}, x_{n,\underline{T}}) = \frac{H(x_{n,\tau:T}, x_{n,1:T-\tau}) + H(x_{m,1:T-\tau}, x_{n,1:T-\tau})}{\sqrt{H(x_{n,\tau:T}, x_{n,1:T-\tau}) \cdot H(x_{m,1:T-\tau}, x_{n,1:T-\tau})}} - \frac{H(x_{m,1:T-\tau}, x_{n,\tau:T}, x_{n,1:T-\tau}) - H(x_{n,1:T-\tau})}{\sqrt{H(x_{n,\tau:T}, x_{n,1:T-\tau}) \cdot H(x_{m,1:T-\tau}, x_{n,1:T-\tau})}} \in [0, 1]\,,$$

where we use $\tau = 1$ as the default lag value.

The corresponding TE matrices denoted as $T_{\underline{N},\underline{N}}$ are generally asymmetric and have diagonal values of 0.

4.4 Convergent Cross Mapping

Convergent cross mapping (CCM) is an inference method for causality between coupled systems developed by Sugihara et al. [1]. It is based on the idea that the attractor of a dynamical system can be reconstructed by shadow manifolds, which are mapped to each other by neighboring states. We refer to Tsonis et al. [9] for a more detailed description of CCM.

4.4.1 Shadow Manifolds

We begin by briefly explaining the mathematical justification of this technique, which is Takens' theorem. It serves as a basis for state-space reconstruction and introduces the concept of delay coordinates and their extensions, the shadow manifolds.

Takens' Theorem

In essence, this theorem states that there exists a smooth map so that the attractor of a dynamical system can be reconstructed by using a finite number of delay coordinates of its individual time series. For a thorough discussion on Takens' theorem we refer to Huke [10]. A slightly simplified but more intuitive expression of Takens' Theorem is given by:

Theorem. *Suppose that a measured time series $x_{n,\underline{T}} = (x_{n,1}, x_{n,2}, ..., x_{n,T})$ lies on a $\mathcal{D}$-dimensional attractor $\mathcal{M}$ of a deterministic dynamical system of d-th order. The starting point obtains an embedding from the recorded data. A convenient, though not unique, representation is achieved by using delay coordinates, for which a delay vector has the following form:*

$$\boldsymbol{m}_{n,t} \equiv \left(x_{n,t}, x_{n,t-\tau}, \ldots, x_{n,t-(\kappa-1)\cdot\tau}\right) \in \mathbb{R}^{\kappa}\,, \tag{4.12}$$

where $\kappa \in \mathbb{N}^{+}$ is the embedding dimension and $\tau \in \mathbb{N}^{+}$ is the time delay. Taken has shown that embeddings with $\kappa > 2d$ will be faithful generically so that there is a smooth map f such that:

$$\begin{aligned} f : \mathbb{R}^{\kappa} &\longrightarrow \mathbb{R} \\ (\boldsymbol{m}_{n,t}) &\longmapsto x_{n,t+1}\,. \end{aligned} \tag{4.13}$$

In essence, this theorem states that the future state of a variable in a dynamical system can be predicted by a finite set of embedded past states, irrespective of other variables in the system. Developing this idea further, the attractor $\mathcal{M}$ of an N-dimensional dynamical system is represented by the shadow manifold of one single system variable $\mathcal{M}_{n,\tau,\kappa}$, which is the set of iterated delay coordinates:

$$\mathcal{M}_n = \begin{pmatrix} \boldsymbol{m}_{n,1+(\kappa-1)\cdot\tau} \\ \boldsymbol{m}_{n,2+(\kappa-1)\cdot\tau} \\ \vdots \\ \boldsymbol{m}_{n,t} \end{pmatrix} = \begin{pmatrix} x_{n,1+(\kappa-1)\cdot\tau} & x_{n,1+(\kappa-2)\cdot\tau} & \cdots & x_{n,1} \\ x_{n,2+(\kappa-1)\cdot\tau} & x_{n,2+(\kappa-2)\cdot\tau} & \cdots & x_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,t} & x_{n,t-\tau} & \cdots & x_{n,t-(\kappa-1)\cdot\tau} \end{pmatrix}, \tag{4.14}$$

where we assume $\kappa > 3$ for illustration purposes. In the following, we will briefly discuss how τ and κ are determined.

Lag Value

We obtain the optimal lag value τ by finding the first local minimum of the MI of the lagged time series. Therefore, for a given time series $x_{n,t}$ we calculate $I(x_{n,\underline{T}}, x_{n,\underline{T}-\tau})$ for different lags τ.

Embedding Dimension

While Takens' theorem theoretically delivers a lower bound for the embedding dimension κ in terms of the order of the dynamical system d, this bound is often unknown. Therefore, Kennel, Brown, and Abarbanel [11] proposed the *False Nearest Neighbor* (FNN) algorithm, which finds the minimal embedding dimension in order to preserve the structure of the attractor $\mathcal{M}$. In essence, the algorithm checks for increasing κ whether neighbors in the original time series remain neighbors in an embedded version. The optimal embedding dimension is found when the fraction of false neighbors falls below a predefined threshold.

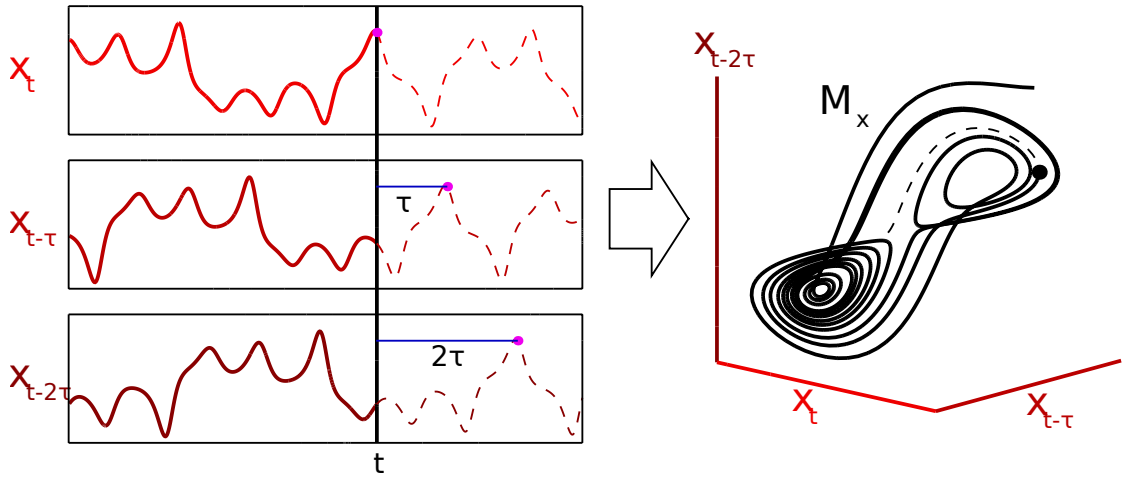

Figure 4.1: Shadow manifold creation within CCM adapted from [12]

4.4.2 Prediction

The intuition behind the CCM is to pretend two time series $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$ belong to the same dynamical system with shadow manifolds $\mathcal{M}_m$ and $\mathcal{M}_n$, respectively. Then, both shadow manifolds are faithful representations of the attractor $\mathcal{M}$ and nearest neighbors in $\mathcal{M}_m$ should identify the time indices of corresponding nearest neighbors in $\mathcal{M}_n$ due to transitivity.

Hence, for a point $\boldsymbol{m}_{m,t} \in \mathcal{M}_m$ we find its $\kappa + 1$ nearest neighbors $\boldsymbol{m}_{m,t_i}$ and denote their time indices $t_i \in (t_1, \ldots, t_{\kappa+1})$ from closest to farthest. We assign a distance-dependent weight u_i to each of the time indices, with closer neighbors leading to a higher value:

$$u_i = \exp\left\{ -\frac{\left\|\boldsymbol{m}_{n,t} - \boldsymbol{m}_{m,t_i}\right\|}{\left\|\boldsymbol{m}_{n,t} - \boldsymbol{m}_{m,t_1}\right\|} \right\}, \tag{4.15}$$

where $\| * \|$ denotes the Euclidean distance.

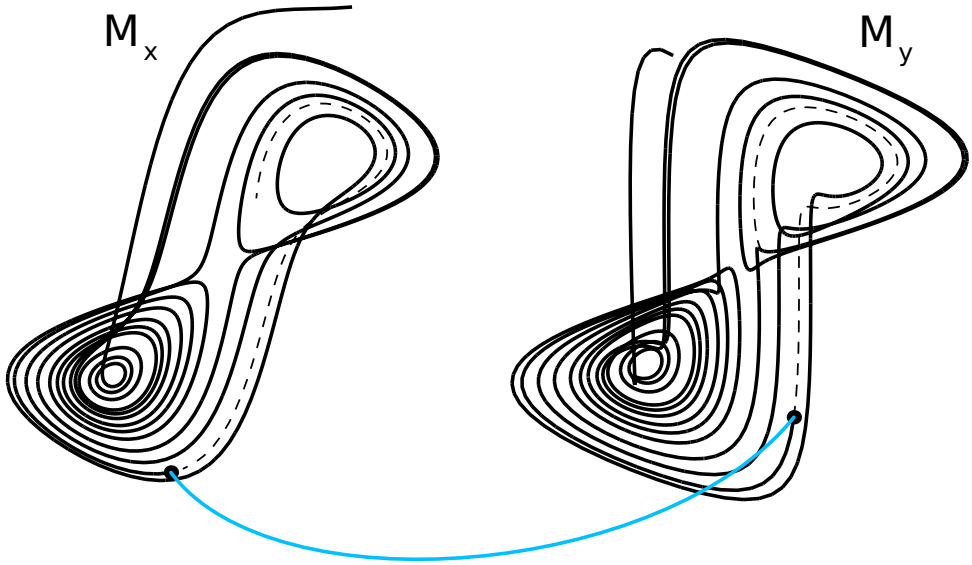

Figure 4.2: Shadow Manifold Cross Mapping within CCM adapted from [12]

.

The weights are normalized via:

$$v_i = \frac{u_i}{\sum_{j=1}^{\kappa+1} u_j} .$$

We can predict a point in the time series $x_{n,\underline{T}}$ by using weighted points in $\mathcal{M}_m$ at time points t_i:

$$\hat{x}_{n,t} = \sum_{i=1}^{\kappa+1} v_i \cdot x_{n,t_i} . \tag{4.16}$$

Thus, under the assumption that $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$ belong to the same dynamical system, the prediction $\hat{x}_{n,t}$ and the actual target value $x_{n,t}$ should be identical. CCM exploits this property and repeats this prediction for a series of L points, which delivers a prediction series $\hat{x}_{n,\underline{L}}$ with its corresponding target series $x_{n,\underline{L}}$.

In order to evaluate the precision of the prediction, we propose several scoring methods. Therefore, we inter alia use different measures from Subsection 3.2.2, which are in general form expressed as $\psi(\hat{x}_{n,\underline{L}}, x_{n,\underline{L}})$:

- The correlation coefficient $\rho(\hat{x}_{n,\underline{L}}, x_{n,\underline{L}})$ between the prediction and the target series is the method used originally by Sugihara et al. [1]. This will be our default method.

- In order to capture nonlinear properties, we use the mutual information $I(\hat{x}_{n,\underline{L}}, x_{n,\underline{L}})$.
- A common measure for prediction evaluations is the coefficient of determination or R^2, which compares the variance explained by the prediction to the total variance:

$$R^2 \equiv 1 - \frac{\sum_{t=1}^{L}(x_{n,t} - \hat{x}_{n,t})^2}{\sum_{t=1}^{L}(x_{n,t} - \mu(x_{n,t}))^2} \in [0,1]\,. \tag{4.17}$$

4.4.3 Scoring

Furthermore, CCM assumes that the more time series points are used for constructing the shadow manifold $\mathcal{M}_m$, the more precise the prediction will be. Thus, we construct $\mathcal{M}_m$ using an increasing number of points, wherefrom we obtain a series of scores. Therefore, we follow the following procedure:

1. Firstly, we split the two time series into a training $x_{m,1:L} \in \mathbb{R}^L$ and test set $x_{m,L+1:T} \in \mathbb{R}^{T-L}$.
2. Subsequently, we divide the training set $x_{m,1:L}$ into W flexible windows as described in Subsection 2.3.2. For each window $w \in \{1, \ldots, W\}$ we calculate shadow manifolds $\mathcal{M}_m$, which increase in density for higher w. In the context of CCM we will refer to the number of windows W as the learning rate η and set the minimum window size to $T_{min} = 10$.
3. Last, for each window we follow the prediction procedure of Subsection 4.4.2, wherefrom we obtain the evolution of scores:

$$\psi_{\underline{W}}(\hat{x}_{n,\underline{L}}, x_{n,\underline{L}}) \in \mathbb{R}^W\,,$$

   which theoretically should converge to the true value of the causal coupling.

We check the convergence of $\psi_{\underline{W}}(\hat{x}_{n,\underline{L}}, x_{n,\underline{L}})$ using the procedure described in Subsection 3.2.1. If the series converges, we take the α-percentile value of the last n_s values in order to smooth the score as illustrated in Subsection 3.2.1. Thus, we define the CCM score between two time series $x_{m,\underline{T}}$ and $x_{n,\underline{T}}$ as:

$$C(x_{m,\underline{T}}, x_{n,\underline{T}}) \equiv \mu^{\alpha}(\psi_{W-n_s:W}(\hat{x}_{n,\underline{L}}, x_{n,\underline{L}}))\,. \tag{4.18}$$

In cases where $\psi_{\underline{W}}(\hat{x}_{n,\underline{L}}, x_{n,\underline{L}})$ does not converge, we set $C(x_{m,\underline{T}}, x_{n,\underline{T}}) = 0$.

In the default case using the Pearson correlation as the scoring method, CCM matrices $C_{\underline{N},\underline{N}}$ have diagonal values of 1.

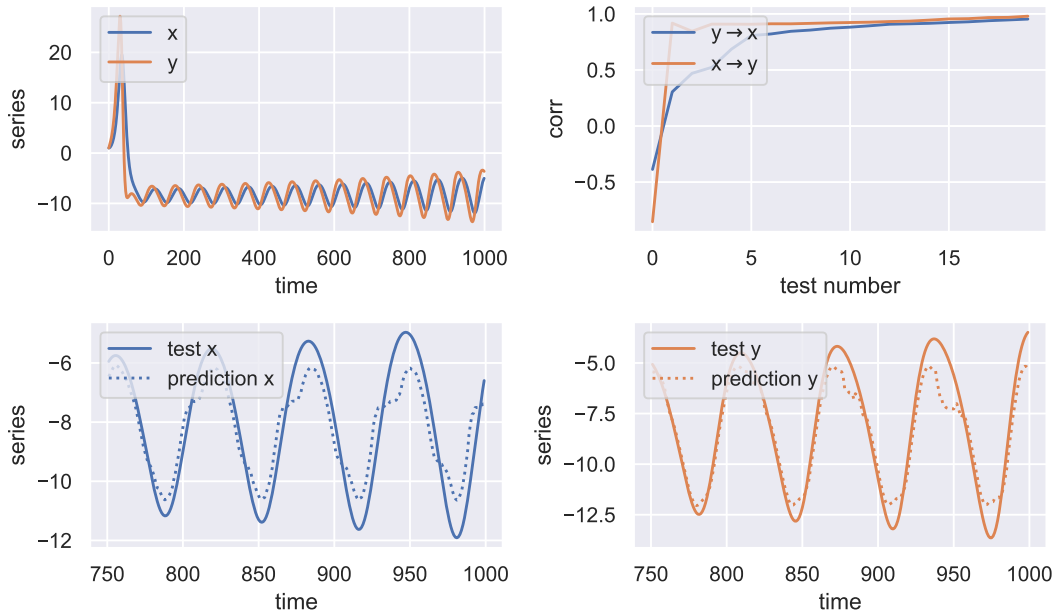

Figure 4.3: CCM inference for the x and y coordinates of the Lorenz system.

4.5 Causal Chains

Inspired by *Markov chains* (MC), we introduce a novel method in order to quantify sub-relations between causally linked state variables. Therefore, we begin with a simplified example of a four-dimensional system with state variables X_1, X_2, X_3, and X_4. Their corresponding causality matrix is denoted as:

$$\psi_{\underline{N},\underline{N}} \equiv \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \psi_{1,3} & \psi_{1,4} \\ \psi_{2,1} & \psi_{2,2} & \psi_{2,3} & \psi_{2,4} \\ \psi_{3,1} & \psi_{3,2} & \psi_{3,3} & \psi_{3,4} \\ \psi_{4,1} & \psi_{4,2} & \psi_{4,3} & \psi_{4,4} \end{pmatrix}, \tag{4.19}$$

where $\psi_{m,n} \geq 0$ denotes the direct causal influence of X_m on X_n for $m, n \in \{1, \ldots, 4\}$. However, $\psi_{1,2}$ for example does not incorporate the full causal influence of X_1 on X_2 since it only captures the direct influence. We identify other potential influences by constructing causal chains:

$$\begin{aligned} &X_1 \xrightarrow{\psi_{1,3}} X_3 \xrightarrow{\psi_{3,2}} X_2 \\ &X_1 \xrightarrow{\psi_{1,4}} X_4 \xrightarrow{\psi_{4,2}} X_2 \\ &X_1 \xrightarrow{\psi_{1,3}} X_3 \xrightarrow{\psi_{3,4}} X_4 \xrightarrow{\psi_{4,2}} X_2 \\ &X_1 \xrightarrow{\psi_{1,4}} X_4 \xrightarrow{\psi_{4,3}} X_3 \xrightarrow{\psi_{3,2}} X_2 \,, \end{aligned} \tag{4.20}$$

wherefore we introduce a set of rules:

- The chain stops when the system variable X_2 is reached for the first time.
- We exclude loops:

$$X_1 \xrightarrow{\psi_{1,3}} \left(X_3 \xrightarrow{\psi_{3,1}} X_1 \xrightarrow{\psi_{1,3}} X_3 \xrightarrow{\psi_{3,1}} \dots \right) \xrightarrow{\psi_{3,2}} X_2 \,,$$

- We exclude recurring system variables:

$$X_1 \xrightarrow{\psi_{1,4}} X_4 \xrightarrow{\psi_{4,3}} X_3 \xrightarrow{\psi_{3,1}} X_1 \xrightarrow{\psi_{1,2}} X_2 \,,$$

  where the system variable X_1 appears twice.

Hence, we refer to the chains in equation 4.20 as causal chains of order $\mathcal{O}(1)$ since every system variable only appears once.

We quantify the causal influence of each chain using the geometric mean over the causality of each link, which in the simplified case yields:

$$\begin{aligned} c_1 &= (\psi_{1,3} \cdot \psi_{3,2})^{\frac{1}{2}} \\ c_2 &= (\psi_{1,4} \cdot \psi_{4,2})^{\frac{1}{2}} \\ c_3 &= (\psi_{1,3} \cdot \psi_{3,4} \cdot \psi_{4,2})^{\frac{1}{3}} \\ c_4 &= (\psi_{1,4} \cdot \psi_{4,3} \cdot \psi_{3,2})^{\frac{1}{3}} \,. \end{aligned} \tag{4.21}$$

The geometric mean ensures that if one causal element in the chain is 0, the total causality for the chain diminishes as well.

Hence, we quantify the total causal influence of X_1 on X_2 as the mean over all possible chains n_c:

$$\Psi_{1,2} \equiv \frac{1}{n_c} \sum_{i=1}^{n_c} c_i \,, \tag{4.22}$$

with $n_c = 4$ in the simplified case.

Extending this principle to an N-dimensional system, for one distinct pair (X_m, X_n) the total number of causal chains is:

$$n_c = \sum_{i=1}^{N-2} \frac{(N-2)!}{(N-i-2)!} \,. \tag{4.23}$$

This can easily be derived by building all possible sub-combinations between the system variables excluding the pair (X_m, X_n).

For all pairs (X_m, X_n), the total number of causal chains is given by:

$$N_c = \binom{N}{2} \cdot n_c = \frac{N!}{2!} \cdot \sum_{i=1}^{N-2} \frac{1}{(N-i-2)!} \,. \tag{4.24}$$

Analogously, we define causal loops for equal pairs (X_n, X_n) to quantify the feedback a state variable X_n receives for a change of one unit.

In this case, the number of chains or loops for one variable is given by:

$$n_l = \sum_{i=1}^{N-1} \frac{(N-1)!}{(N-i-1)!} \,. \tag{4.25}$$

In combination with causal chains, we can construct causality matrices by quantifying the causal link between X_m and X_n using $\Psi_{m,n}$ instead of $\psi_{m,n}$. The resulting matrix is denoted as $\Psi_{\underline{N},\underline{N}}$ with diagonal values quantifying the causal loop feedback. We use the trace of the matrix in order to quantify the system causality:

$$\zeta\big(\psi_{\underline{N},\underline{N}}\big) \equiv \frac{1}{N} \operatorname{tr}\big(\Psi_{\underline{N},\underline{N}}\big) \,. \tag{4.26}$$

The attentive reader might have noticed that this definition violates the integrity condition specified in Section 4.1. However, we can solve this problem by setting the diagonal values to a constant value when constructing causality matrices.

Note that for a system of $N = 3$, this measure reduces to the geometric norm described in Subsection 3.2.3.

Chapter 5

Causality Driver Analysis

In this chapter we apply our model to identify and quantify linear and nonlinear drivers of causality in synthetic time series systems. To do so, we firstly compare the different inference techniques methodically using the coupled logistic system as a benchmark before applying them to the Lorenz attractor.

5.1 Measure Comparison

We compare GC, TE, and CCM regarding their causality evolution and robustness regarding window size, causal coupling, and noise. In this Subsection we conduct our analysis on the coupled logistic model specified in Subsection 2.2.1. The default parameters are chosen and the simulation length is set at $T = 5000$.

5.1.1 Evolution

We begin our analysis by calculating the causality evolutions over fixed-size sliding windows with size $w = 1000$ and sliding delta $\delta T = 200$. We increase the robustness of our results by using $K = 15$ realizations of FT surrogates.

Pearson Correlation

In order to conduct a validity check of our model, we begin by evaluating the Pearson correlation. As discussed in Subsection 3.2.2, we expect the evolution of the correlation to be identical for the original and surrogate data. Furthermore, the cross measure evolutions are theoretically supposed to have a similar shape.

We observe in Figure 5.1 that our expectations are met and conclude that our surrogate generation is correct.

Due to the large size of our window, we do not explicitly measure the anti-correlation phases arising from mirage correlations mentioned in Subsection 2.2.1. However, we observe periods of different correlation strength.

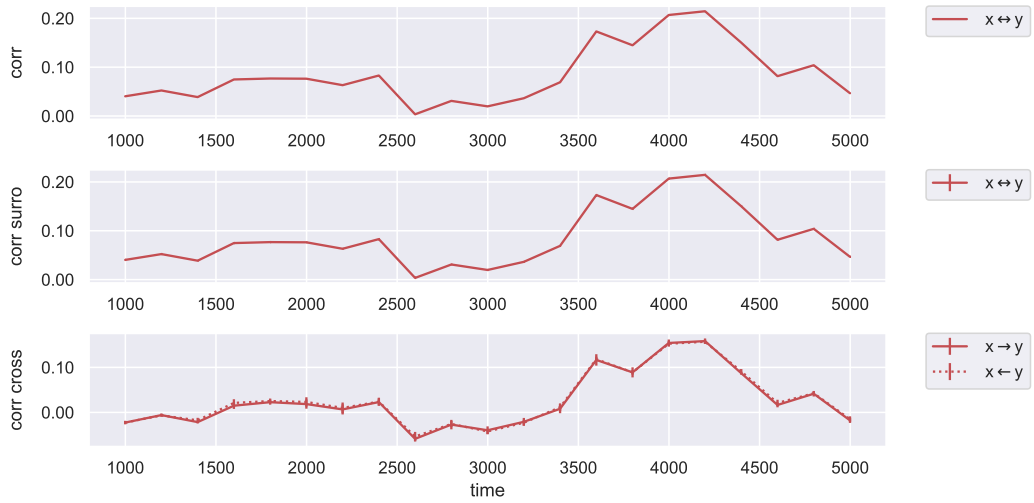

Figure 5.1: Pearson correlation in the coupled logistic system for fixed-size windows.

Mutual Information

In order to ensure the correctness of our PDF estimation via histograms, we check whether the evolution of the surrogate MI has the same shape as the Pearson correlation.

For a bin size of $n_b = 75$ and a bin range of $r_b = (0.005, 0.995)$ we observe a high similarity between the surrogate MI in Figure 5.2 and the Pearson correlation in Figure 5.1. This is numerically confirmed by a nested correlation of $\rho = 0.93$.

Furthermore we find a high nested correlation of $\rho = 0.89$ between the surrogate MI and the cross MI from y to x. This indicates that the nonlinear part in x has no contribution to the MI. This also applies to the other direction, for which the nested correlation has a lower value of $\rho = 0.81$.

We do however observe an increase of original MI between timesteps $t = 2500$ and $t = 3500$ which is not present in the surrogate or cross data. We suggest the following interpretation: since the cross MI in both directions have very similar evolutions as the surrogate MI, for which the surrogatization is performed on both time series instead of one, this means that the nonlinear part in one time series has no strong

effect on MI. However, by comparing the evolutions to the original MI we see a significant difference. Hence, we can conclude that only the interaction between the nonlinearities of both time series affects MI while the nonlinearity of one single time series seems to be irrelevant.

Checking the additive nature between the MI evolutions did not generate findings that would warrant further studies.

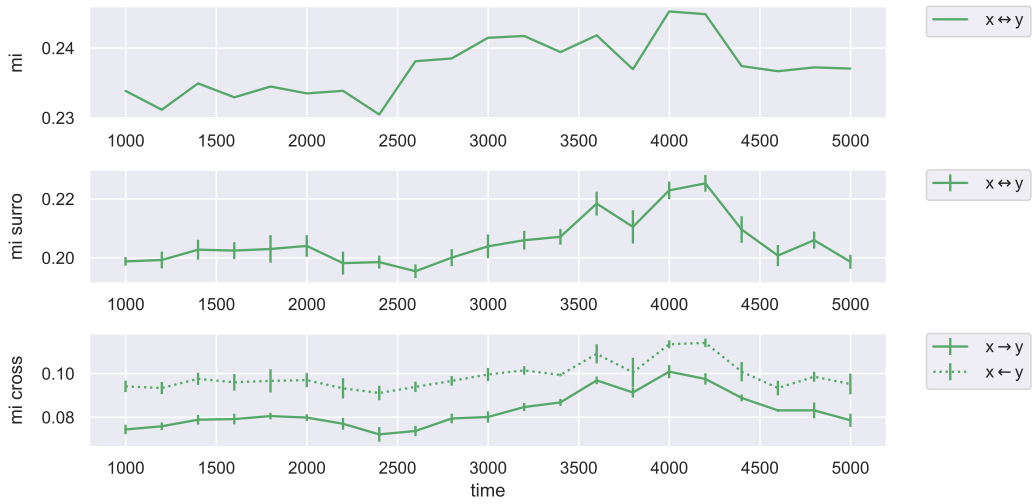

Figure 5.2: Mutual information in the coupled logistic system for fixed-size windows.

Granger Causality

As discussed in Section 4.2, GC measures solely the linear causality between time series. We observe this in Figure 5.3 by comparing the evolutions of the original and surrogate GC in both directions. We find them to be identical backed by a nested correlation of $\rho = 0.99$. Generally, the data indicates a higher linear causality from the x to the y coordinate and practically no causality the other way around. This is inconsistent with the causal coupling chosen for the coupled logistic system. We will later find evidence that in this particular system, the causality from y to x is mainly driven by nonlinearities and hence cannot be detected by GC.

Between time steps $t = 3400$ and $t = 4700$ we observe an increase in cross causality from x to y, which is also slightly present between $t = 3800$ and $t = 4400$ for the original and surrogate GC. We find that these artifacts arise from coherence phases of mirage correlations, which are detected as causality by GC.

Henceforth, we will mostly drop GC from our analysis since it offers no benefits for finding nonlinear causality drivers.

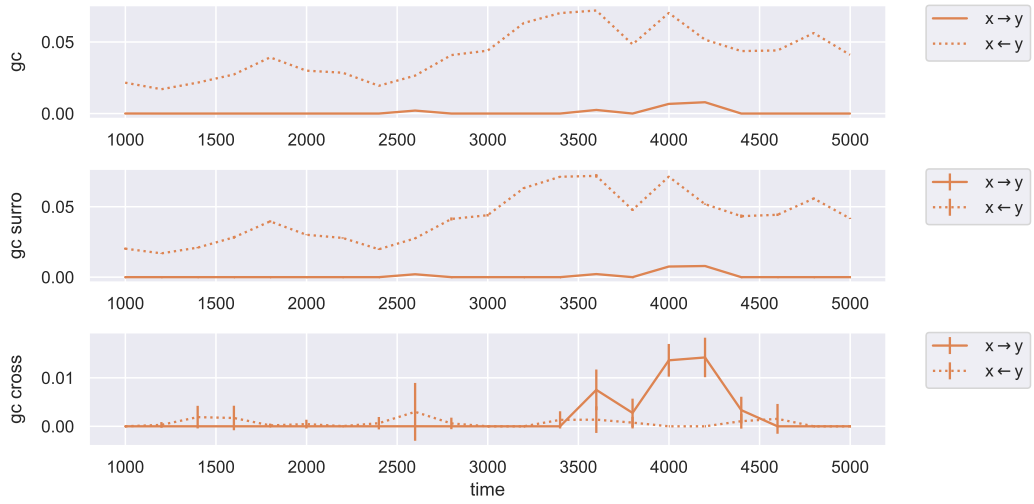

Figure 5.3: GC in the coupled logistic system for fixed-size windows.

Transfer Entropy

In comparison to GC, TE incorporates both linear and nonlinear properties of causality in its inference. Analogously to MI, the bin size chosen is $n_b = 75$ and the bin range is set at $r_b = (0.005, 0.995)$. Figure 5.4 illustrates the causality evolutions detected by TE, which in contrast to GC, indicate a higher causality from x to y for both original and surrogate data. This is consistent with the causal coupling of the system. The cross causality further suggests that the causality from y to x is mainly driven by linear properties in y.

We observe less fluctuation of the TE evolutions for the original data compared to the surrogate data. This can be interpreted as meaning that the total causality consistently keeps the same level while the linear and nonlinear contributions fluctuate.

Even though GC theoretically is supposed to be the linear special case of TE, we do not find that the surrogate TE evolutions match with GC. Due to the difficulty with scoring GC, this was not to be expected. However, this calibration could be subject to further research in order to validate findings from analyses with TE.

Generally, we must note that the consistency of TE across original and surrogate data needs further research since TE appears to be very sensitive to the width of the distributions. Especially for the PDF estimation using histograms, we find that a sufficient resolution within the core ranges of the distributions is necessary for a robust estimation irrespective of the bin size. Preprocessing techniques such as rank-ordered remapping or rescaling described in Subsections 2.3.3 and 2.3.4 did not yield significant improvements.

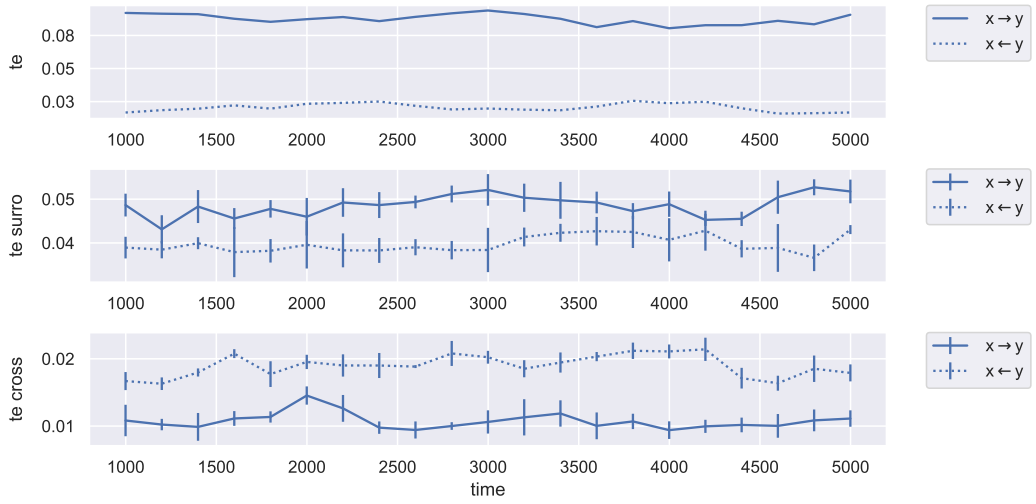

Figure 5.4: TE in the coupled logistic system for fixed-size windows.

Convergent Cross Mapping

In comparison to PDF-dependent measures, CCM has a consistent normalization due to its scoring method, wherefore we use the absolute value of the Pearson correlation $|\rho|$. Furthermore, for our analysis we choose the lag value $\tau = 1$, embedding dimension $\kappa = 8$, learning rate $\eta = 0.05$, and smoothing constant $n_s = 5$.

As we observed for TE, the CCM causality from x to y is stronger for both original and surrogate data. Generally, Figure 5.5 indicates that the causality for the original is higher than the surrogate time series for both directions. Hence, we conclude that the causality in the coupled logistic system is mainly driven by nonlinear properties.

However, we observe that the evolution of the cross causality from y to x is identical to the original causality with a nested correlation of $\rho = 0.99$. At first glance, this suggests that the causality from y to x is either entirely driven by the time series x or the linear part of y according to the following reasoning: the only difference between the original and cross causality is the surrogatization of the time series y. Since this procedure has no apparent effect on the causality, we can conclude that either y has no effect on the causality at all or that only the linear part of y contributes.

Due to this special case, we look at the original and surrogate time series x and y in Figure 5.6. While the original and surrogate series of x differ recognizably, this does not apply to y. The highly periodic structure leads to a peaked power spectrum and hence the series is not significantly affected by the phase randomization.

One could therefore deduce that the original and surrogate time series of y are identical and thus the causality from y to x must be entirely driven by the time

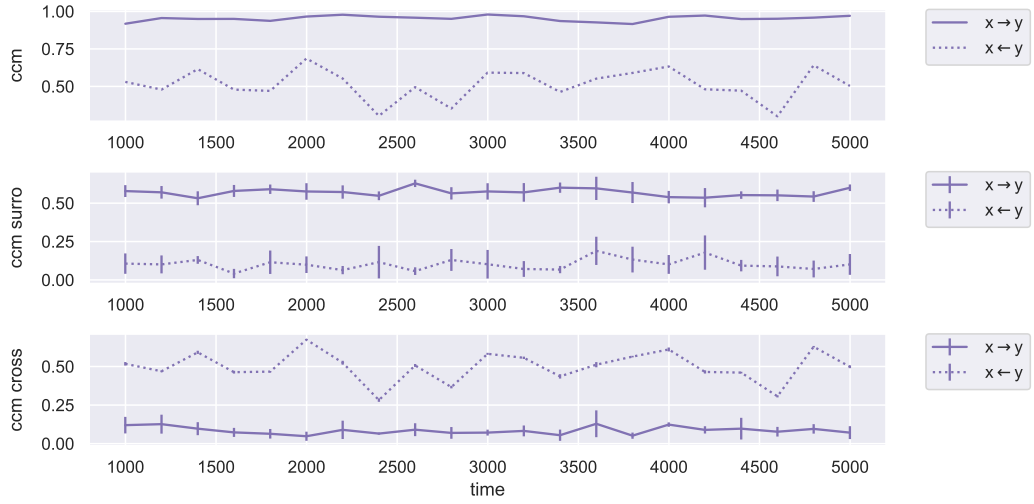

Figure 5.5: CCM in the coupled logistic system for fixed-size windows.

series x. However, in this case the cross causality from x to y should be identical to the surrogate causality in this direction. Since this does not apply, we can rule out that the causality is entirely encoded in the time series x.

This leads us to the following reasoning: the only difference between the surrogate causality and the cross causality from y to x is the surrogatization of the time series x. Since the surrogate causality almost vanishes in contrast to its corresponding cross measure, this indicates that the causality is majorly driven by the nonlinearity in x.

We observe that the cross causality from x to y vanishes in comparison to its corresponding original. Hence, we can conclude that the causality in this direction is also mainly driven by x. However, the surrogate causality remains at a constant level between the original and cross causality. We refer back to our analysis of MI, where we suggested that interactions of the respective linear and nonlinear parts of the time series play an important role. This could be an area for further research.

Generally, our analysis of the causality evolutions leads us to the conclusion that nonlinearity is an important driver of causality. We do however find that attributing causality to linear or nonlinearity leaves room for interpretation. This leads us to defining different methods in order to capture different aspects of nonlinear causality present in the system.

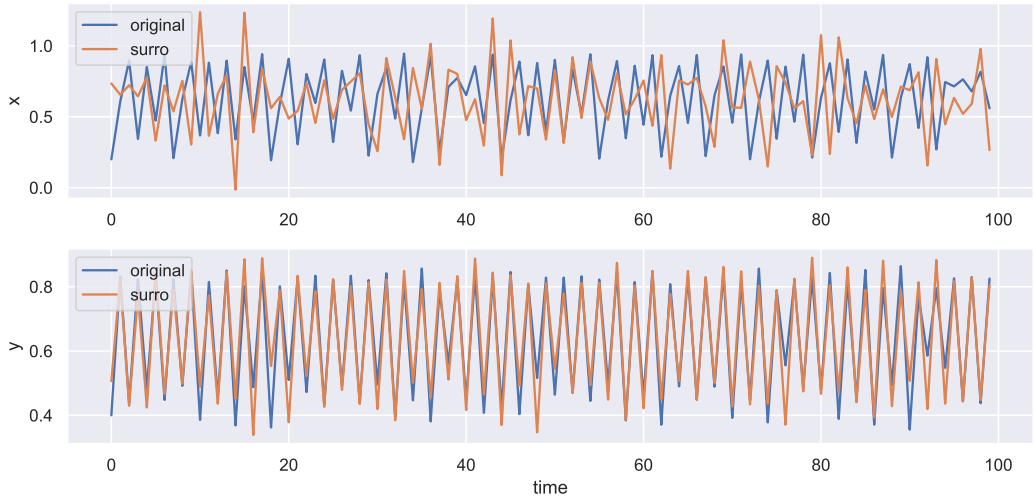

Figure 5.6: Original and surrogate time series snippet of the coupled logistic system.

5.1.2 Nonlinearity

As specified in Subsection 3.1.4, we quantify the nonlinear causality by performing several transformations using the original, surrogate, and cross causality. Since GC does not measure nonlinearities, we analyze the nonlinear MI along with TE and CCM. We present a selection of four nonlinearity measures in this section.

For the maximum nonlinearity, TE detects no nonlinearity from the y to the x coordinate, while CCM detects high nonlinearity for the other direction. This is illustrated in Figure 5.7.

Considering that the measures should be consistent, we look at the relative nonlinearity in figure 5.8. We observe for both TE and CCM that the nonlinear causality from y to x is higher than for the other direction. This is in agreement with the reasoning from the previous section.

The relative cross nonlinearity illustrated in Figure 5.9 shows consistent results across all measures with higher nonlinearity from x to y, but no nonlinear TE and CCM for the other direction.

For the relative cross surrogate nonlinearity depicted in figure 5.10 we observe very high nonlinearity for TE and CCM in both directions with higher values from x to y.

In conclusion, it is difficult to find a universally valid nonlinearity measure since nonlinear causality can be understood in several ways. We leave it up to the reader to decide which definition of nonlinearity is appropriate for the respective research problem.

Figure 5.7: Maximum nonlinearity in the coupled logistic system.

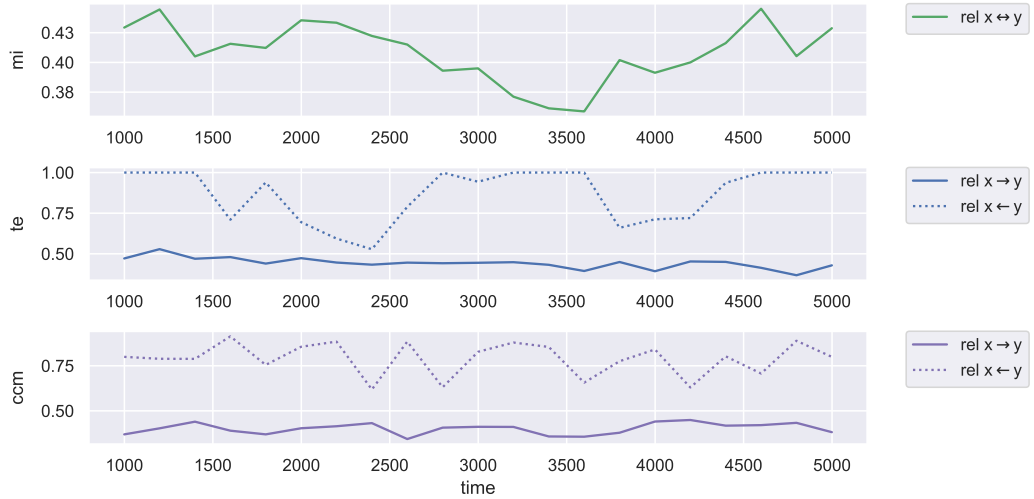

Figure 5.8: Relative nonlinearity in the coupled logistic system.

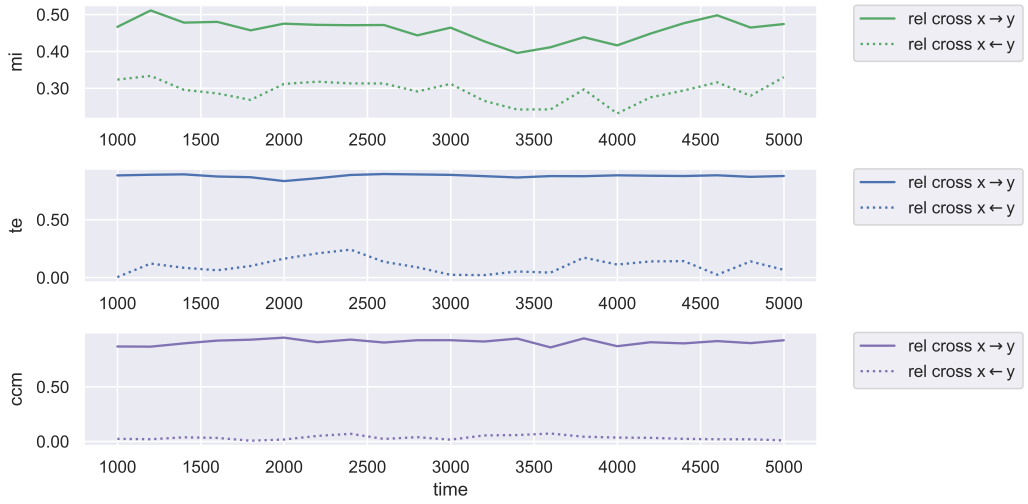

Figure 5.9: Relative cross nonlinearity in the coupled logistic system.

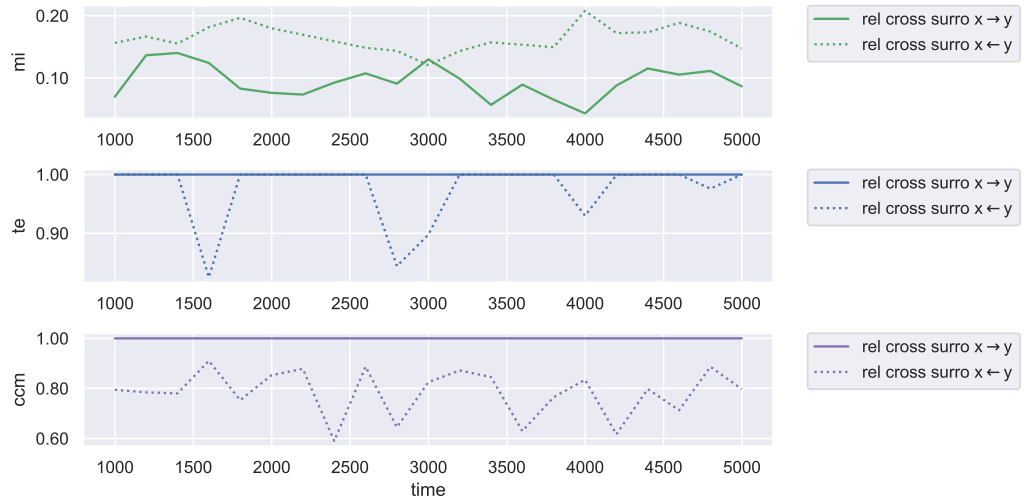

Figure 5.10: Relative cross surrogate nonlinearity in the coupled logistic system.

5.1.3 Window Size

In order to check the robustness of the causality inference methods for varying window sizes, we analyze the same system using flexible windows. We slice the time series into $W = 20$ windows with minimum and maximum sizes between $T_{min} = 100$ and $T_{max} = 10000$. The parameters for the individual methods remain generally unchanged.

Granger Causality

For GC we observe a slowly increasing level of causality from y to x and no causality the other way around. After a window size of $T_w = 8000$, GC remains at a constant level. The evolution of the original and surrogate GC is almost identical, except for the first window. This can be attributed to the estimation error due to the small window size. From the errorbar we see that the original GC lies within the standard deviation of the surrogate GC.

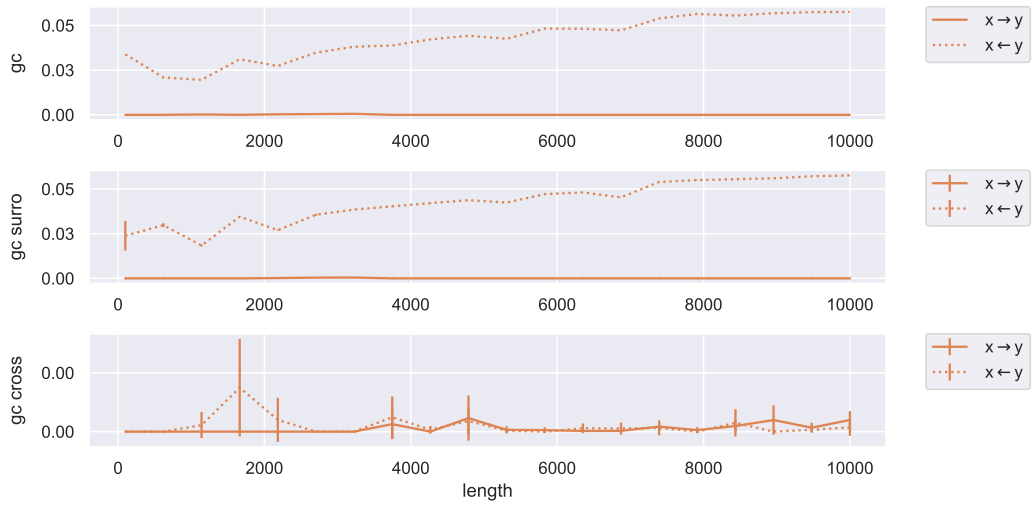

Figure 5.11: GC in the coupled logistic system for flexible-sized windows.

Transfer Entropy

For TE we observe a steady but slow decline for the x to y original causality. This result is expected due to the binning for the PDF histogram estimation, which is not optimized for every window size. Generally, we must note that for all PDF dependent measures the binning for the histogram estimation can heavily distort the results.

The surrogate causality shows a similar decline with a change in causality strengths at a window size of approximately $T_w = 500$. In contrast to the causality evolution, the causality from y to x is stronger than the other direction. Furthermore, we observe a pronounced zigzag behavior, which could arise from added phases of higher causality.

The cross causality from x to y shows a steeper decline while the other direction slowly vanishes in a similar zigzag pattern.

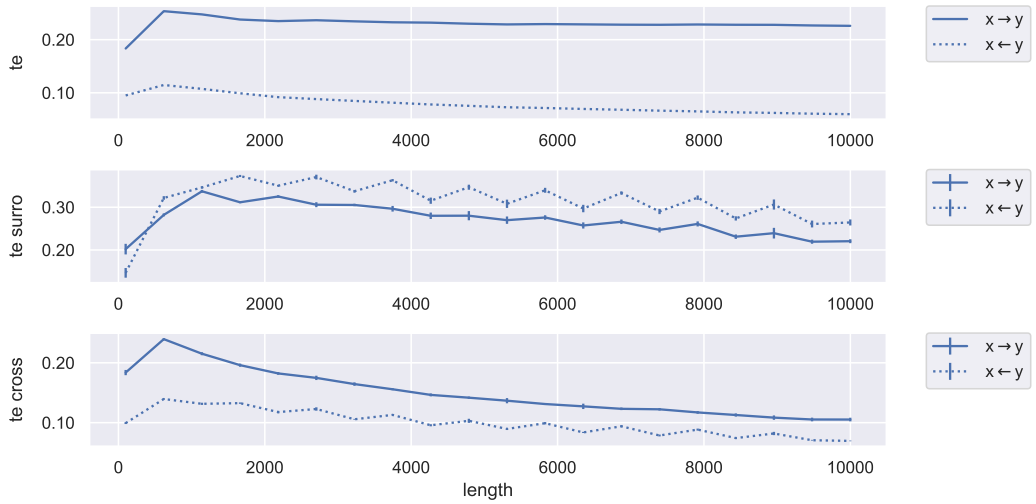

Figure 5.12: TE in the coupled logistic system for flexible-sized windows.

Convergent Cross Mapping

While the causality from x to y remains at a constant level, we see a slow increase in the other direction for larger window sizes. This can be explained by the higher attractor density, which improves the CCM estimation as explained in Subsection 4.4.2. An equilibrium is approximately reached after a window size of $T_w = 7000$.

We observe a distinct zigzag pattern at a steady average level for the surrogate causality from x to y, while the other direction quickly vanishes after a window size of $T_w > 500$.

After a change of causality strengths at $T_w = 500$, the cross causality from y to x surpasses the other direction, which quickly vanishes thereafter. We observe a a slowly increasing zigzag pattern with a stronger variation.

Generally, we find that CCM detects the relations between the causality levels consistently with regards to the evolutions from Subsection 5.1.1.

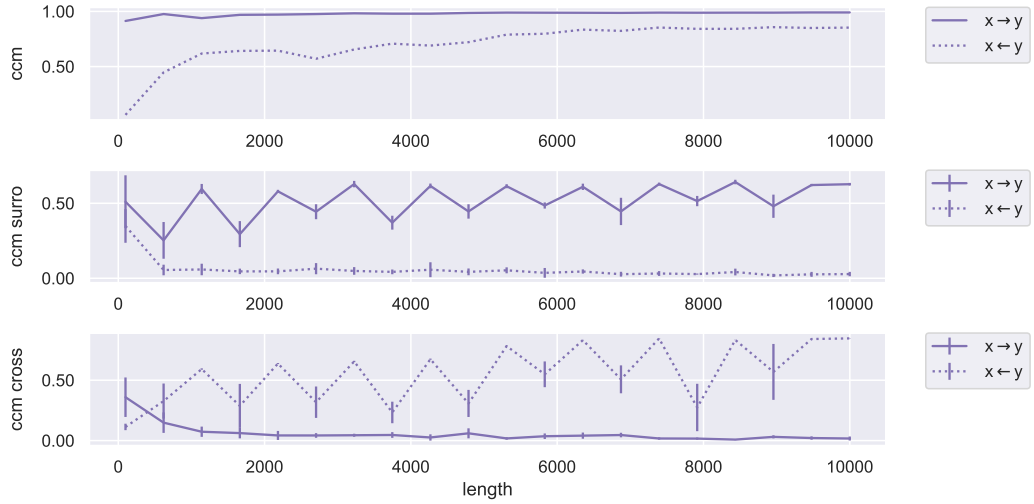

Figure 5.13: CCM in the coupled logistic system for flexible-sized windows.

5.1.4 Causal Coupling

Another point of comparison is the correct detection of the causal couplings. Analogous to Sugihara et al. [1], we simulate the coupled logistic system for $T = 400$ and vary the coupling strength for both $\beta_{x,y}$ and $\beta_{y,x}$ in a grid using the ranges $(0, 0.4)$. This delivers matrices, which we can illustrate using heatmaps. Note that we label the axes differently since we switched the indices in order to adhere to our model notation.

Transfer Entropy

Due to the smaller window size in comparison to Subsection 5.1.1, we choose a bin size of $n_b = 30$ and keep the bin range at $r_b = (0.005, 0.995)$.

In Figure 5.14 we observe an increasing gradient from bottom to top, which is consistent with the coupling strength in this direction. We do however observe an area where almost no TE is measured. This can be explained by higher mirage correlations for these parameter configurations. In a thorough analysis we find that this area is dependent on the number coherent phases and can be eliminated by choosing a larger simulation length.

The same phenomenon can be observed for the other causality direction as illustrated in Figure 5.15. However, in general the gradient of the matrix indicates that the causal coupling is measured correctly. Due to the almost perfect symmetry of the coupled logistic system, the unidirectional matrices are almost transposed versions of each other.

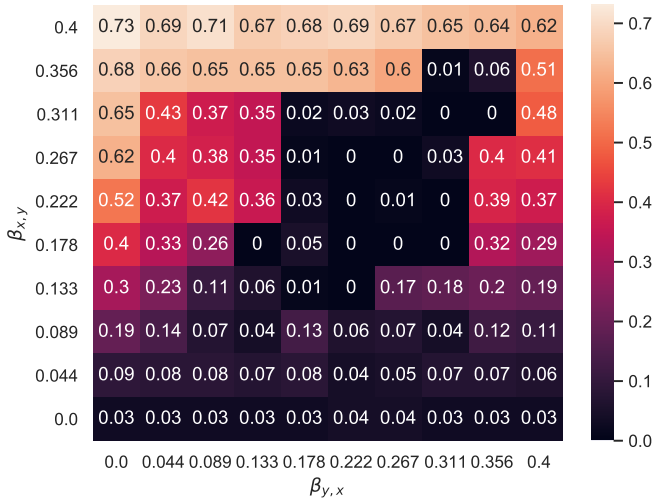

Figure 5.14: TE from x to y for varying causal couplings.

Similarly to Sugihara et al. [1], we substract the matrices from each other to get a bidirectional causality matrix. As expected, we observe a symmetric gradient along the diagonal, which is presented in Figure 5.16.

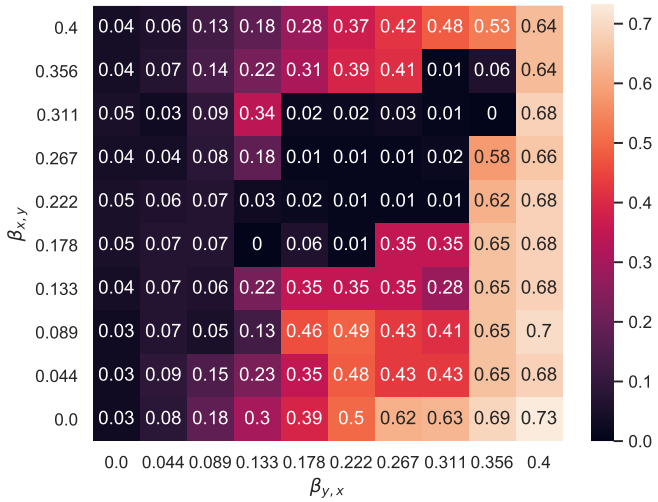

Figure 5.15: TE from y to x for varying causal couplings.

For surrogate data we see no obvious gradient but rather an even coloring as shown in Figure 5.17. This indicates that the causality structure vanishes if nonlinear properties are destroyed. Furthermore, the absolute values are generally much lower than for the original data. This is further evidence for the hypothesis that causality is mainly driven by nonlinearity.

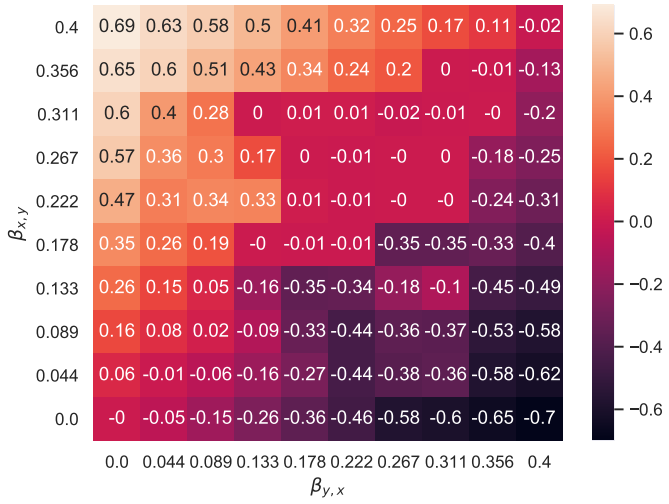

Figure 5.16: Bidirectional TE for varying causal couplings.

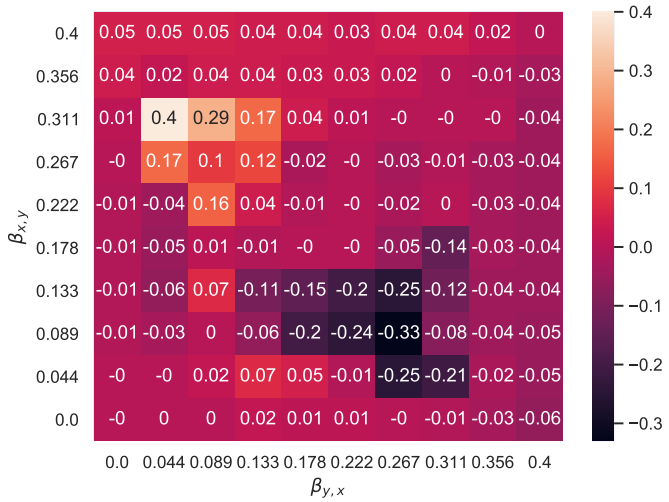

Figure 5.17: Bidirectional surrogate TE for varying causal couplings.

Convergent Cross Mapping

Analogously, we perform the parameter grid analysis for CCM. In comparison to the parameter setting in Subsection 5.1.1, we choose the Pearson correlation ρ for scoring and an embedding dimension of $\kappa = 5$.

Firstly, we compute the bidirectional matrix and compare it to Sugihara et al. [1] for validation. We obtain the same gradient and range of the inferred causality as depicted in Figure 5.18. As illustrated in the benchmark, we also obtain outliers and no even gradient which is symmetric along the diagonal. While this issue is not addressed by the authors, our analysis indicates that the outliers result from failed causality inferences in the unidirectional matrices. This can be observed in Figures 5.20 and 5.21. They show that no causality is detected for certain parameter configurations.

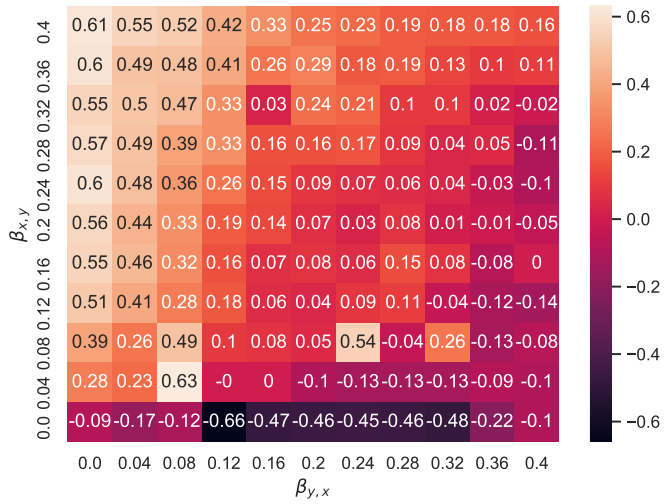

Figure 5.18: Bidirectional CCM for varying causal couplings.

The surrogate bidirectional matrix is shown in Figure 5.19, where we also observe a gradient but with a much lower value range. Thus, we can conclude that a large part of the causality is driven by nonlinear properties. As hinted above, the outliers for the bidirectional surrogate bidirectional matrix are located at the same parameter configurations as for the original data.

Generally, we observe correct gradients for the unidirectional matrices depicted in Figures 5.20 and 5.21. However, for the causality direction from x to y we additionally see a symmetry across the diagonal and low levels of causality in the last column. This is caused by the chaotic nature of the system, which is enhanced if causality in both directions is strongly present.

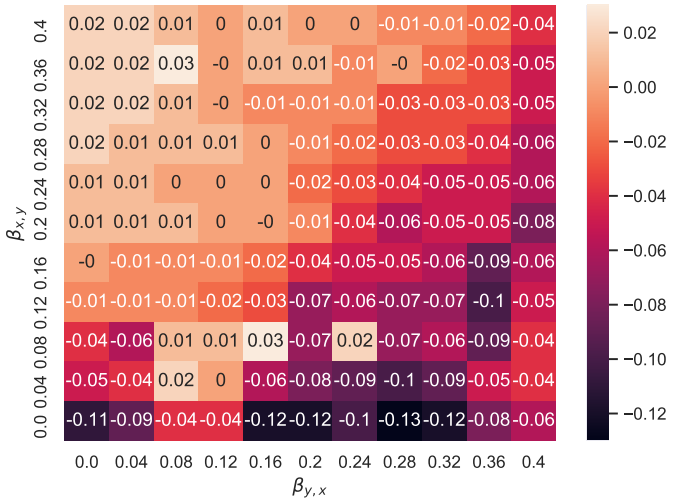

Figure 5.19: Bidirectional surrogate CCM for varying causal couplings.

Furthermore, in contrast to TE, the matrices are not transposed versions of each other. The gradient for the y to x coupling seems consistent up until the last column, which shows significantly lower values. This result is surprising due to the symmetry of the coupled logistic system. In contrast to the other direction, the first column shows significantly higher values. Additionally, the causality values in the last column are also inconsistent with the couplings.

We suspect that Sugihara et al. [1] purposely only illustrated the bidirectional matrix since the subtraction covers up the outliers. Further analyses with longer simulation lengths and other parameter configurations lead to similar results. We conclude that CCM is a good technique to detect whether causality is present between two time series. However, at least for the coupled logistic system, CCM can not be used for detecting the actual causal coupling strengths.

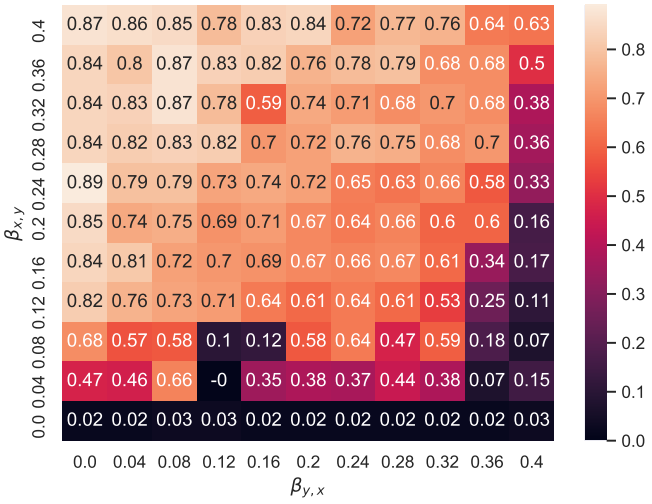

Figure 5.20: CCM from x to y for varying causal couplings.

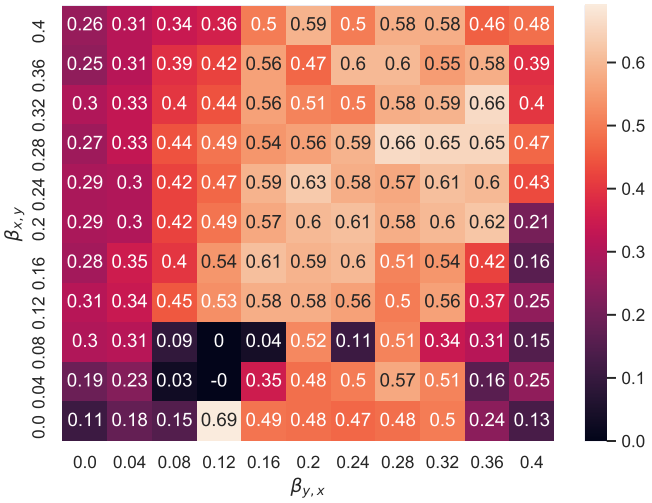

Figure 5.21: CCM from y to x for varying causal couplings.

5.1.5 Noise

In order to ensure our model to be applicable to real-world data, we check the causality inference methods for robustness with regards to noise. Therefore, we add different degrees of Gaussian white noise to the time series. Hence, the time series are transformed to:

$$\begin{aligned} x_t &\longmapsto x_t + \epsilon_{t,x} \cdot W_{t,x} \\ y_t &\longmapsto y_t + \epsilon_{t,y} \cdot W_{t,y} \,, \end{aligned} \tag{5.1}$$

where $W_{t,x}$ and $W_{t,x}$ are Wiener processes and we vary $\epsilon_{t,x}$ and $\epsilon_{t,y}$ between $[0, 1]$ respectively.

Transfer Entropy

For TE from x to y we generally see a strong decline when noise is added. Furthermore, as illustrated in Figure 5.22 the matrix is not symmetric across the diagonal. This means that there is a difference whether noise is added to the driver or target time series. We find the same pattern for higher resolutions of the grid and smaller noise degrees.

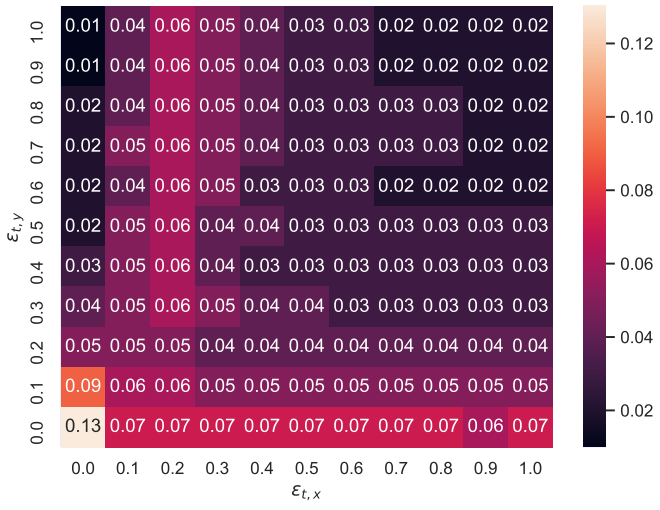

Figure 5.22: TE from x to y for varying noise strengths.

The same break in asymmetry is observed for the direction y to x in Figure 5.23. Interestingly, a higher causality is detected in the second row.

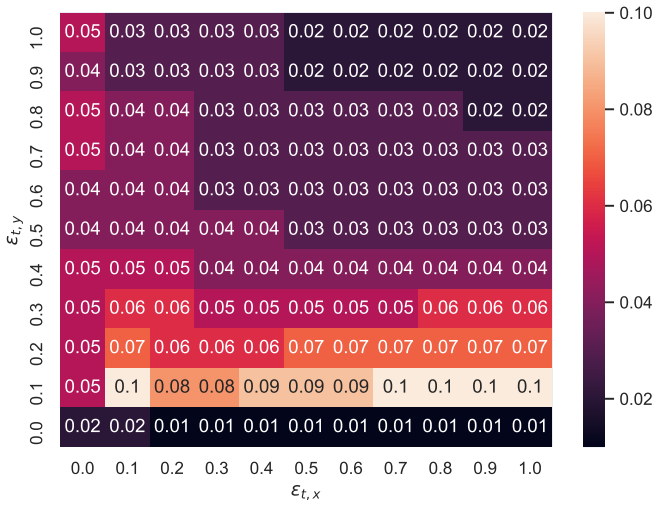

Figure 5.23: TE from y to x for varying noise strengths.

As illustrated in Subsection 2.2.4, we test TE on a random dummy system. Figure 5.24 depicts the bivariate causality for the x and y coordinate of the random Gaussian system. We observe that overall high levels of TE are detected and additionally, find a high nested correlation of $\rho = 0.87$ between the original and cross evolutions.

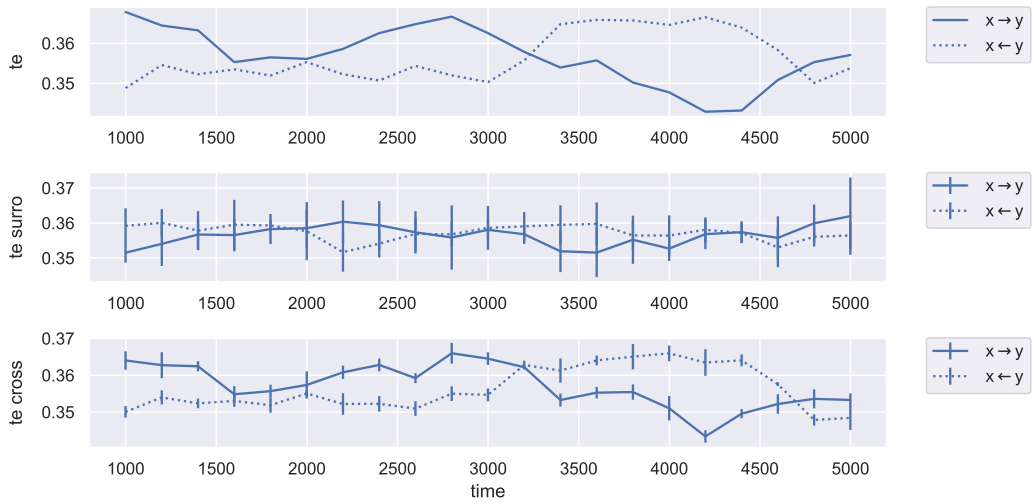

Figure 5.24: TE between x and y in the random Gaussian system.

Convergent Cross Mapping

Analogously for CCM, we find that the noise matrices are more symmetrical across the diagonal and transposed versions of each other. In Figure 5.25 we see for the direction x to y that the causality decreases for weak noise but increases again if stronger degrees are added.

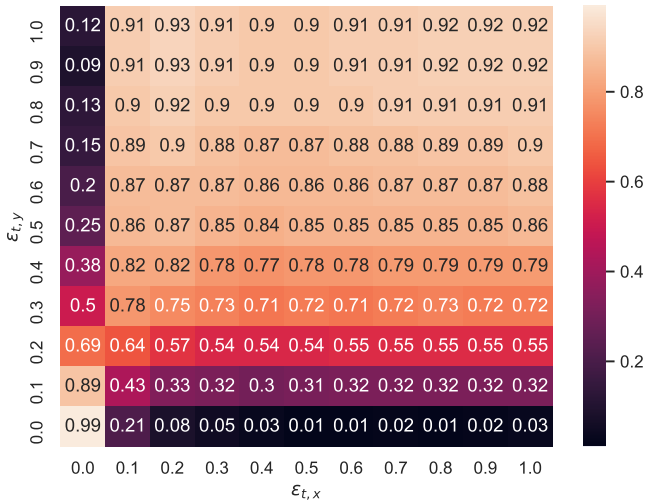

Figure 5.25: CCM from x to y for varying noise strengths.

The same effect can be observed for the y to x direction in Figure 5.26, where the causality even surpasses the level of the unaltered time series.

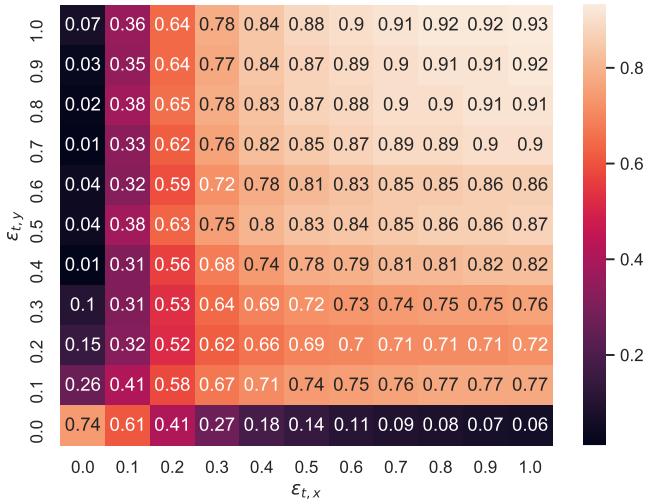

Figure 5.26: CCM from y to x for varying noise strengths.

The high causality values for strong degrees of noise could arise from structural issues with the CCM inference method. Hence, as a backtest we calculate CCM on the random Gaussian system and illustrate the x and y coordinates in Figure 5.27. For original, surrogate, and cross causality we observe fluctuating and generally low values detected by CCM.

At this point, we leave it to further research to find the reasons for the false detection of causality for random time series.

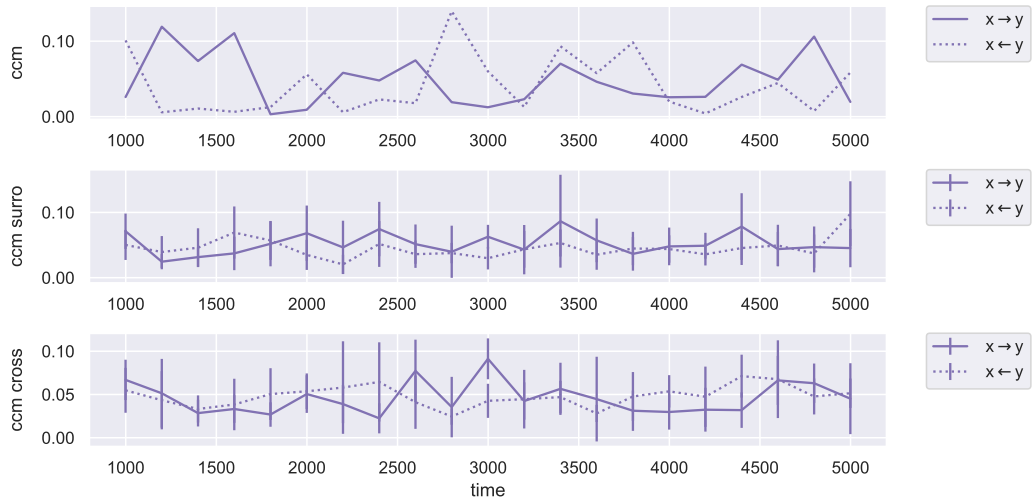

Figure 5.27: CCM between x and y in the random Gaussian system.

5.2 Lorenz System

Due to its popularity we choose the Lorenz attractor defined in Subsection 2.2.2 for our analysis. The default parameter settings are chosen and the simulation length of the system is $T = 50000$ timesteps.

5.2.1 Bivariate

We begin by analyzing the bivariate causality evolution using both TE and CCM. Therefore, we divide the time series into fixed-size sliding windows with size $w = 5000$ and sliding delta $\delta T = 1000$.

Transfer Entropy

For all three coordinate pairs we observe rather constant evolutions for the original data. Furthermore, Figures 5.28, 5.29, and 5.30 confirm that the surrogate causality is lower than the original causality while keeping the causality directions intact.

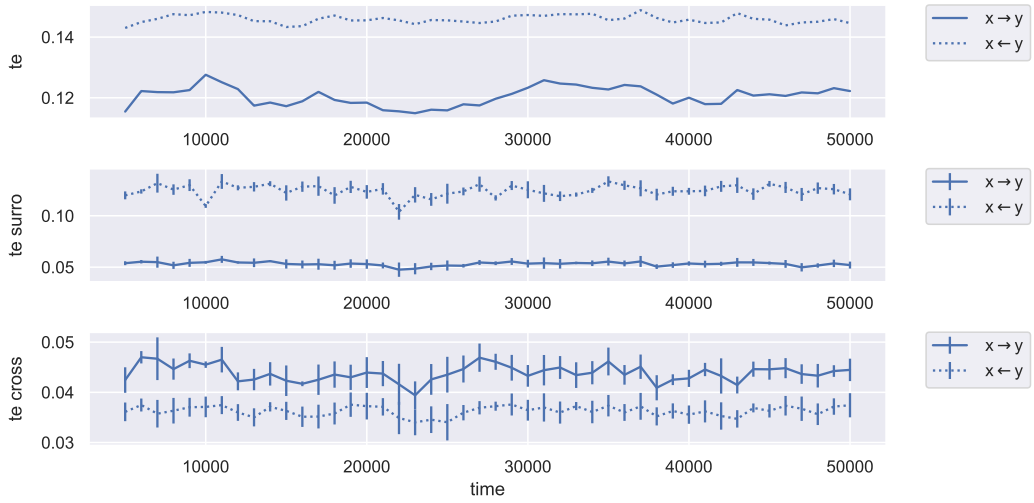

Figure 5.28: TE between x and y in the Lorenz system.

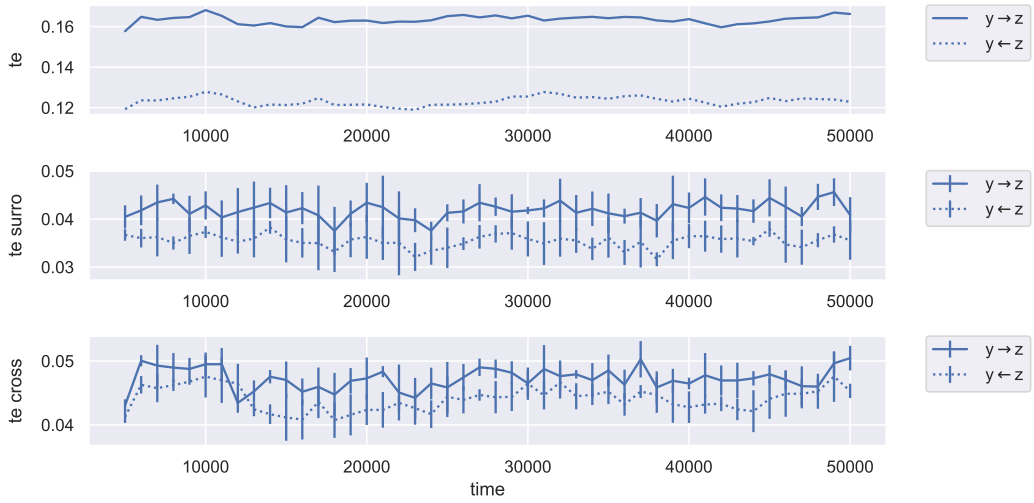

Figure 5.29: TE between y and z in the Lorenz system.

Interestingly, it seems that the causality between the y and z pair is almost identical to x and z, which is confirmed by an average nested correlation of $\rho = 0.91$.

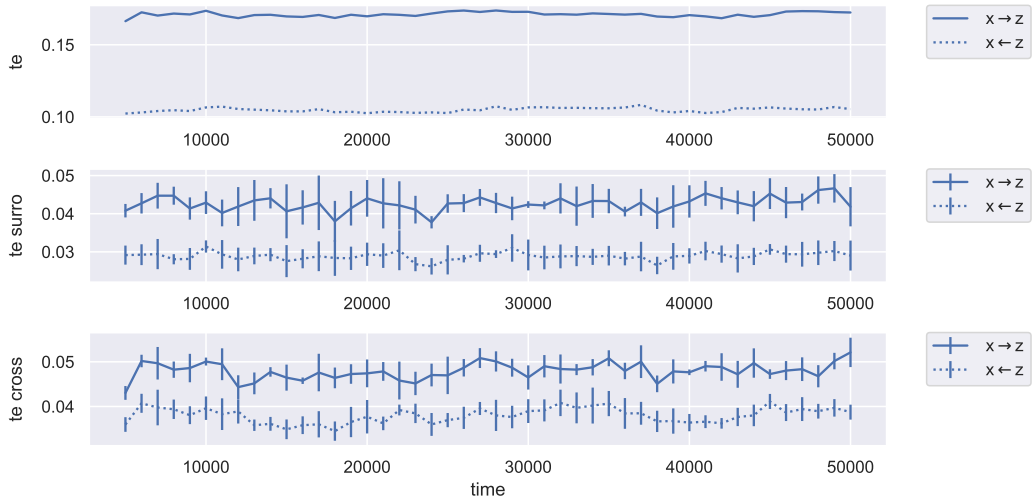

Figure 5.30: TE between x and z in the Lorenz system.

Convergent Cross Mapping

For CCM we find larger differences between the coordinate pairs. For x and y we observe very high causality levels for both original and surrogate data in both directions. Figure 5.31 shows slightly higher levels of causality for the original than the surrogate data. However, the difference is practically negligible.

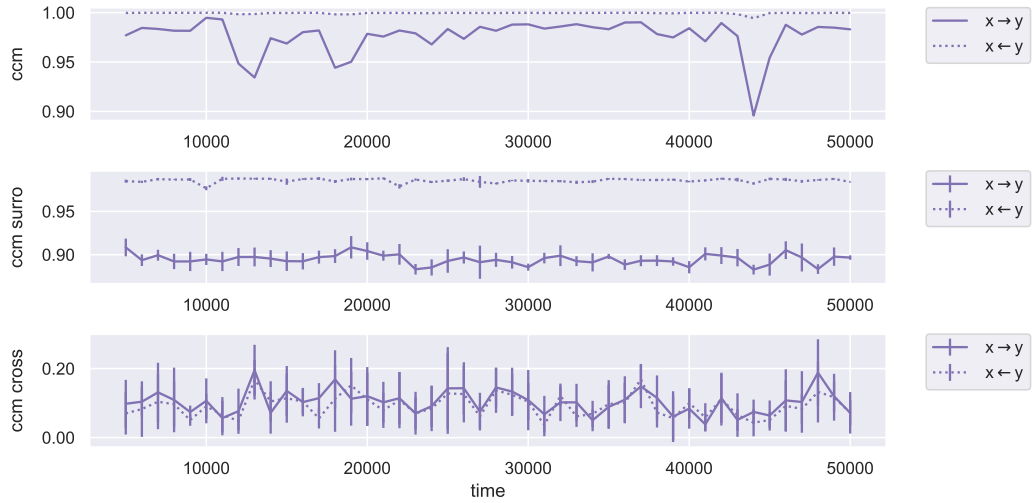

Figure 5.31: CCM between x and y in the Lorenz system.

For the y and z pair illustrated in Figure 5.32, we observe very high causality from z to y and fluctuating medium levels for the other direction. In comparison, the surrogate causality is low for both directions.

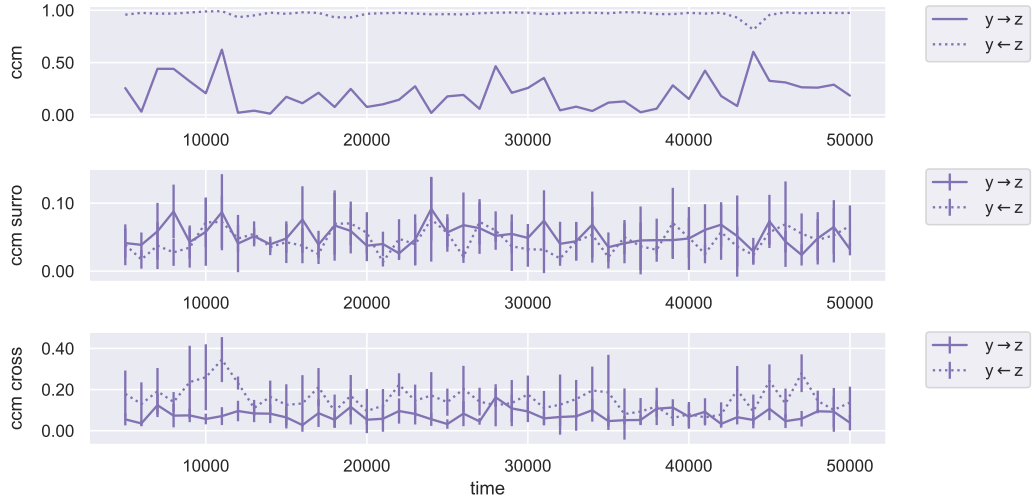

Figure 5.32: CCM between x and y in the Lorenz system.

Analogously to TE, we find almost identical evolutions for the x and z pair depicted in Figure 5.33. This is confirmed by average nested correlations of approximately $\rho = 0.93$

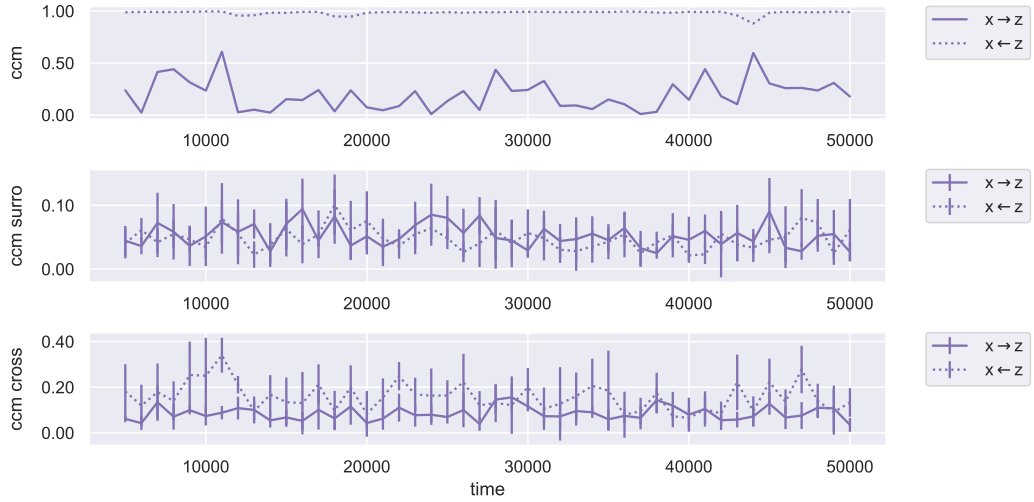

Figure 5.33: CCM between x and z in the Lorenz system.

5.2.2 System Analysis

In order to quantify the system causality of of the Lorenz system, we hereby present a selection of three system measures calculated on MI, TE, and CCM. Recall Subsections 3.1.1 and 4.5 for their respective definitions.

We observe for the absolute norm illustrated in Figure 5.34 and the Frobenius norm in Figure 5.35 that the system measures remain constant over time, even though we saw significant deviations in the bivariate evolutions in Subsection 5.2.1.

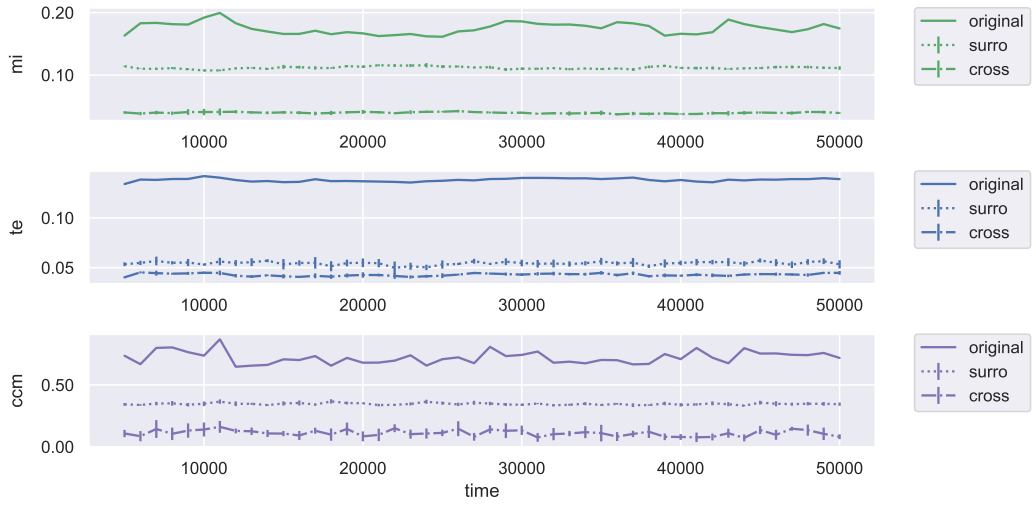

Figure 5.34: Absolute norm measures in the Lorenz system.

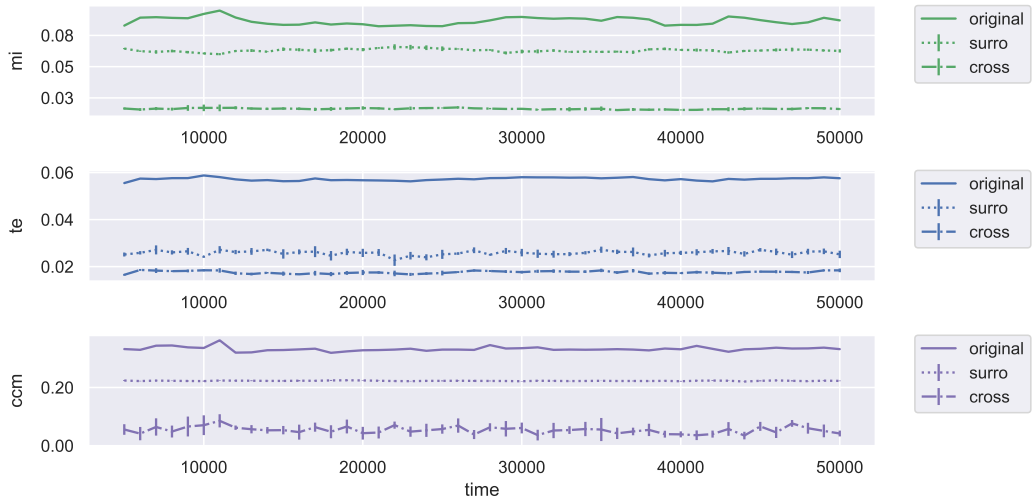

Figure 5.35: Frobenius norm measures in the Lorenz system.

In comparison, we see larger fluctuations of CCM for the geometric chains in Figure 5.36, which is expected due their definition.

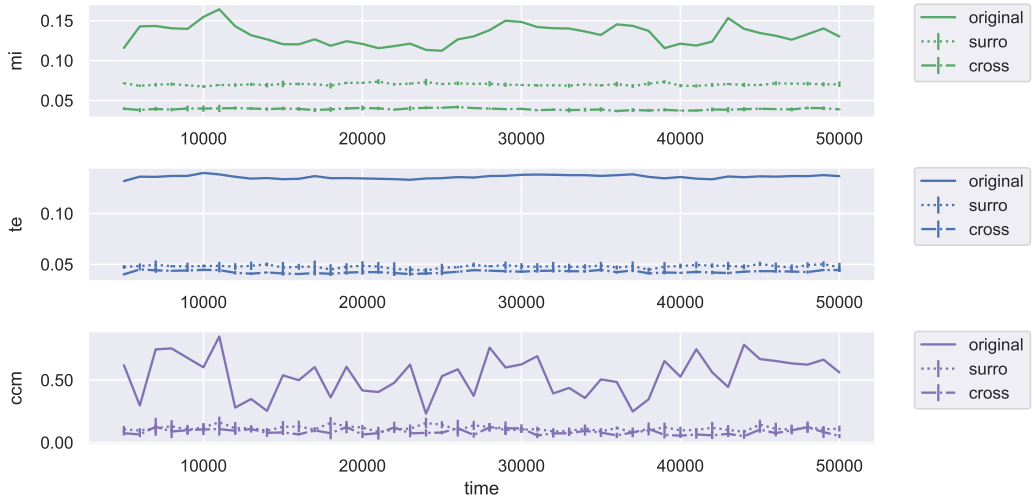

Figure 5.36: Geometric chain causality in the Lorenz system.

Overall, we observe that the measure levels for the original data is higher than for the surrogate and cross data. Thus, we can conclude that the causality in the Lorenz system is mainly driven by nonlinearities.

5.2.3 Decomposition

In order to get a clearer structure of the causality, we decompose the TE and CCM of the system into its bivariate contributions.

Transfer Entropy

The decomposition of the TE shows an almost equal distribution from all causality directions over time. However, we again must note the possible inconsistencies due to the binning issue for PDF dependent measures.

For surrogate data we see that the x and y pair has the largest contribution to the system causality. As indicated in Figure 5.38, especially the y to x causality seems to be almost unaffected by the surrogatization and is the largest contributor to the system causality.

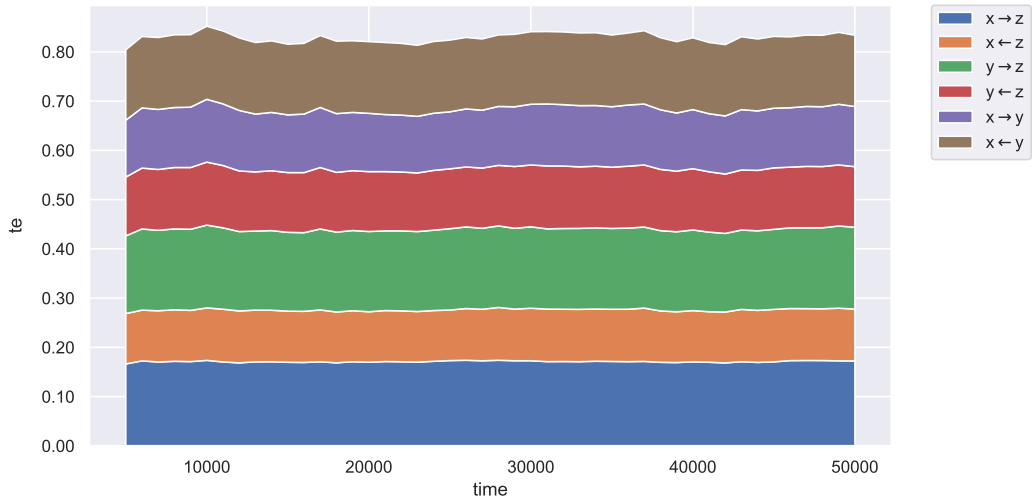

Figure 5.37: TE decomposition in the Lorenz system.

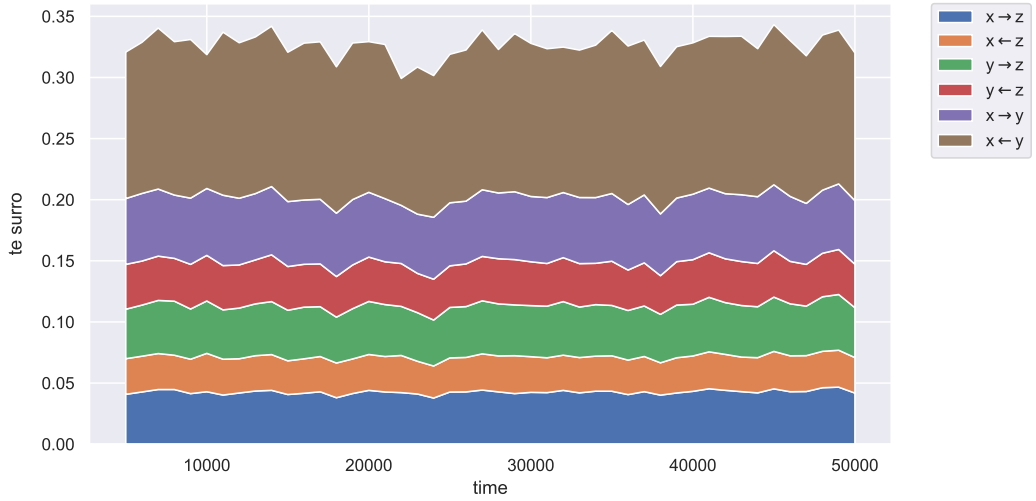

Figure 5.38: Surrogate TE decomposition in the Lorenz system.

Convergent cross Mapping

We see a more uneven decomposition for CCM in Figure 5.39, where, except for the directions x to z and y to z, all directions have similarly large contributions to the system causality.

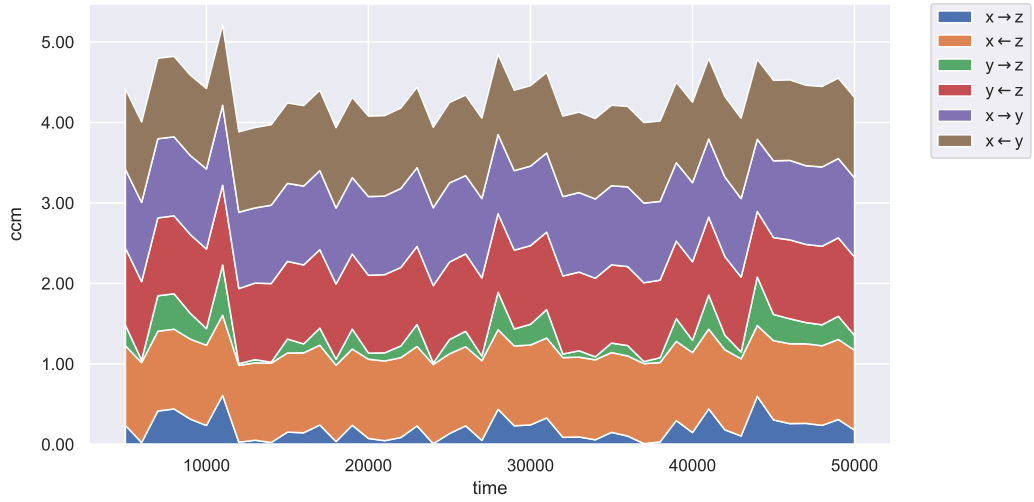

Figure 5.39: CCM decomposition in the Lorenz system.

For surrogate data illustrated in Figure 5.40, we see a completely different decomposition. While the x and y pair dominates the system causality, the other pairs are almost completely diminished.

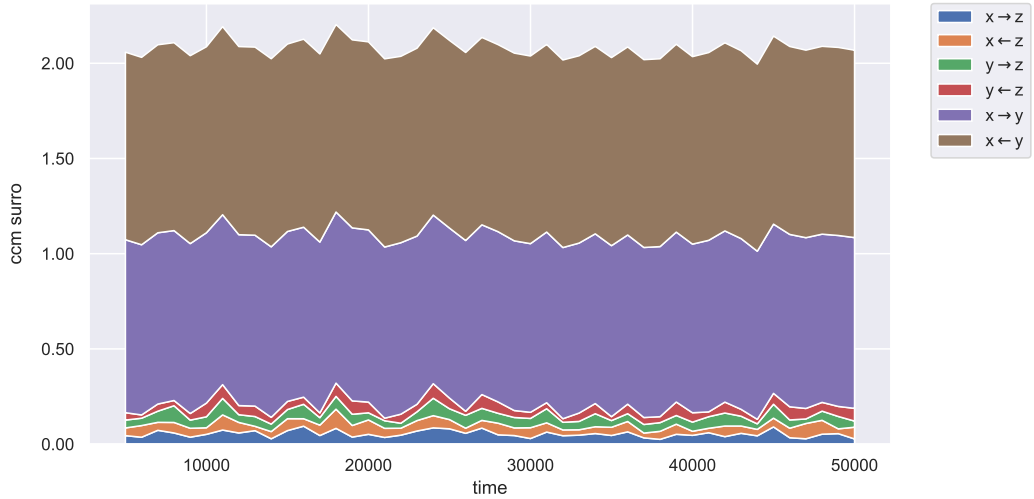

Figure 5.40: TE in the Lorenz system.

Generally, we observed that the x and y pair seems to be unaffected by the surrogatization. We find that this is caused by the underlying time series of the Lorenz attractor.

In Figure 5.41, we see that the x and y coordinate have almost identical evolutions, which explains the high causality levels detected between them.

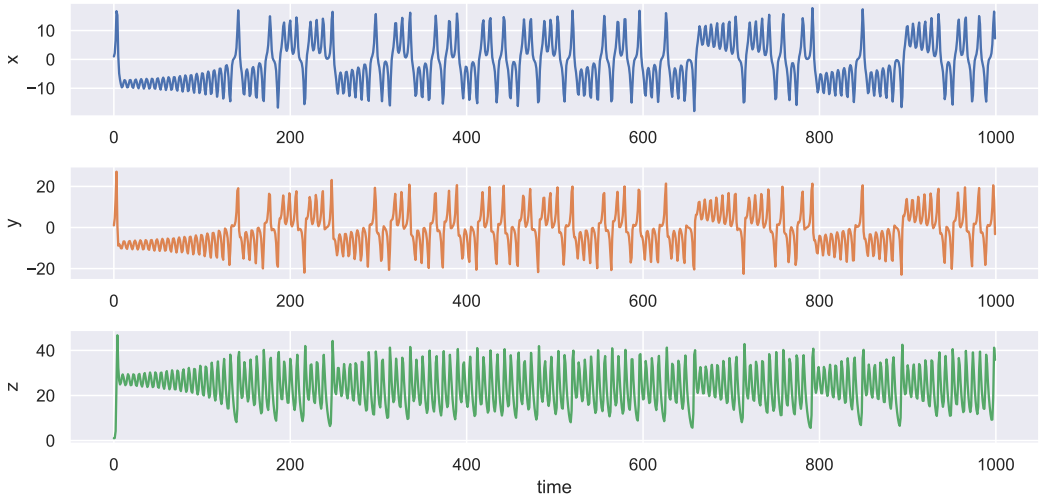

Figure 5.41: Time series of the Lorenz system.

This leads to almost identical surrogate time series as depicted in Figure 5.42, which is an explanation for the high causality between x and y for both original and surrogate data.

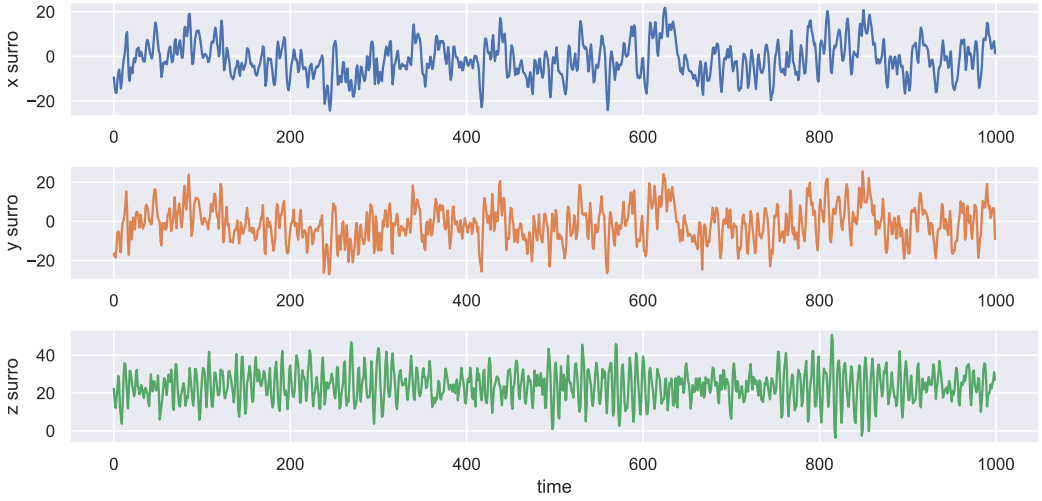

Figure 5.42: Surrogate time series of the Lorenz system.

5.2.4 Asymmetry

In order to quantify the asymmetry of the causality in the Lorenz system, we hereby present two measures which were defined in Subsection 3.1.1. Additionally to TE and CCM, we analyze MI as a backtest since its original and surrogate matrices are perfectly symmetric as per definition.

The absolute asymmetry is depicted in Figure 5.43, where we see a higher asymmetry of the original data in comparison to the surrogate data for both TE and CCM. As designed, the original and surrogate MI show no asymmetry.

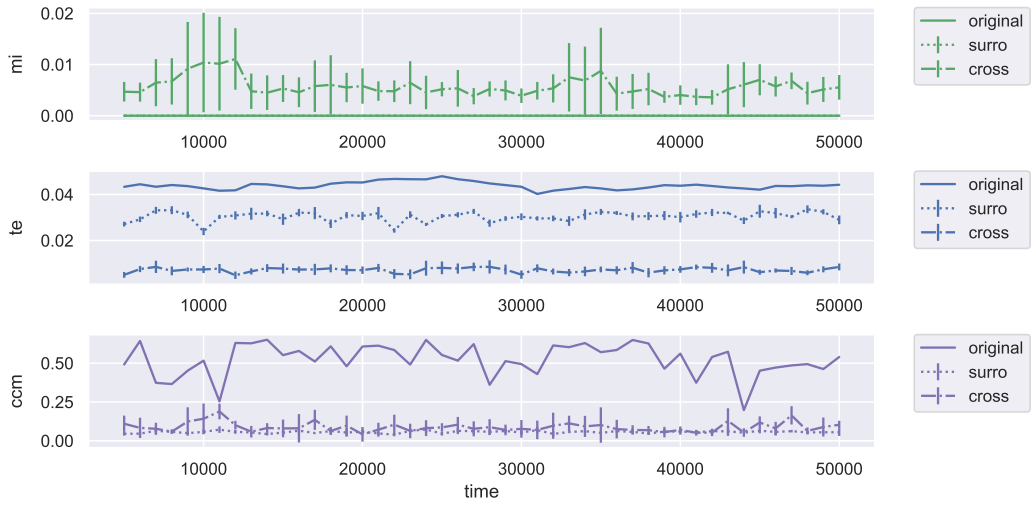

Figure 5.43: Absolute asymmetry in the Lorenz system.

Since the absolute asymmetry is not normalized, the original asymmetry is usually higher than the surrogate asymmetry if causality is driven by nonlinearities. Thus, Figure 5.44 depicts the fractional asymmetry which shows higher surrogate asymmetry for TE but lower surrogate asymmetry for CCM. However, as discussed, the TE results could be distorted due to flawed binning.

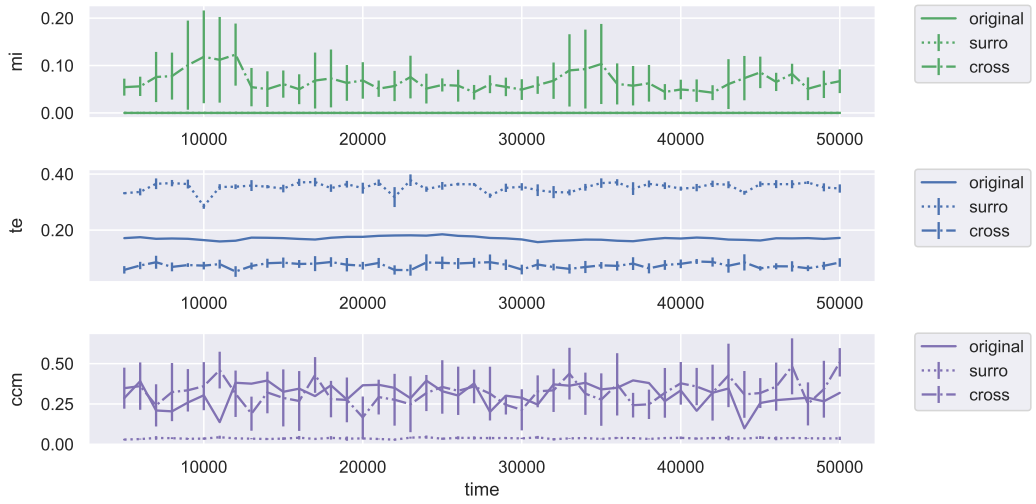

Figure 5.44: Fractional asymmetry in the Lorenz system.

Taking the analyses of other systems into account, we presume that the asymmetry of the causality in nonlinear systems is caused by nonlinear properties. We suggest this topic to be subject to further research.

5.2.5 Attractor

In order to analyze the causality structure within the attractor, we choose smaller non-overlapping windows of size of $T_w = 1000$ and perform the same analysis as in the previous sections. We expect the causality inference to sometimes fail for CCM since it relies on the density of points in the attractor. Thus, this section only fulfills the purpose of detecting underlying structures of causality without contradicting the results from the previous sections.

Transfer Entropy

In comparison to the evolution from Subsection 5.2.1, we observe higher symmetries between the evolutions of the x and y pair as depicted in Figure 5.45. This applies to the original, surrogate, and cross causality. Furthermore, we see identical evolutions for the original and cross TE, which is confirmed by a nested correlation of $\rho = 0.95$.

This phenomenon is analogously present for the y and z pair illustrated in Figure 5.46. We see similar evolutions for both directions of the original data, which are again almost identical to the cross evolutions.

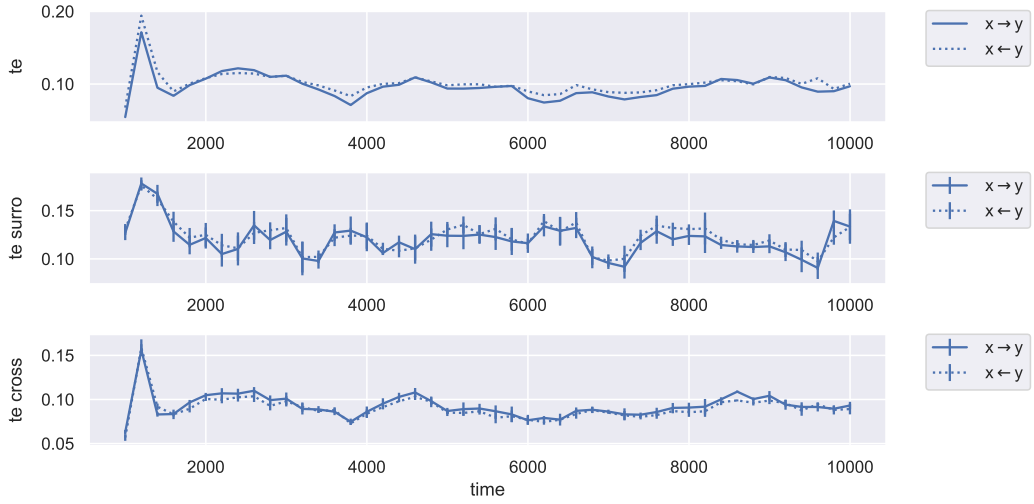

Figure 5.45: TE between x and y for short timeframes in the Lorenz system.

Furthermore, as discussed in Subsection 5.2.1, our proposition that the causality structure between z and the other two coordinates is practically identical is confirmed. In Figure 5.47 we observe the same evolutions for the y and z pair with an average nested correlation of $\rho = 0.96$.

This indicates that the causality measured by TE in short timeframes is almost entirely driven by linear properties. This contradicts our results for large timeframes in Subsection 5.2.1. However, as mentioned above, the short timeframes cannot be seen as being representative for the system.

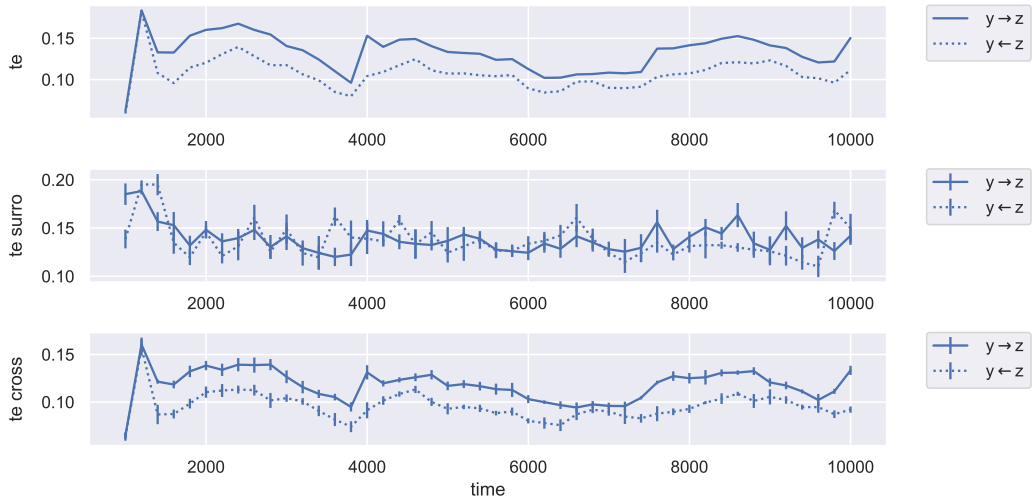

Figure 5.46: TE between y and z for short timeframes in the Lorenz system.

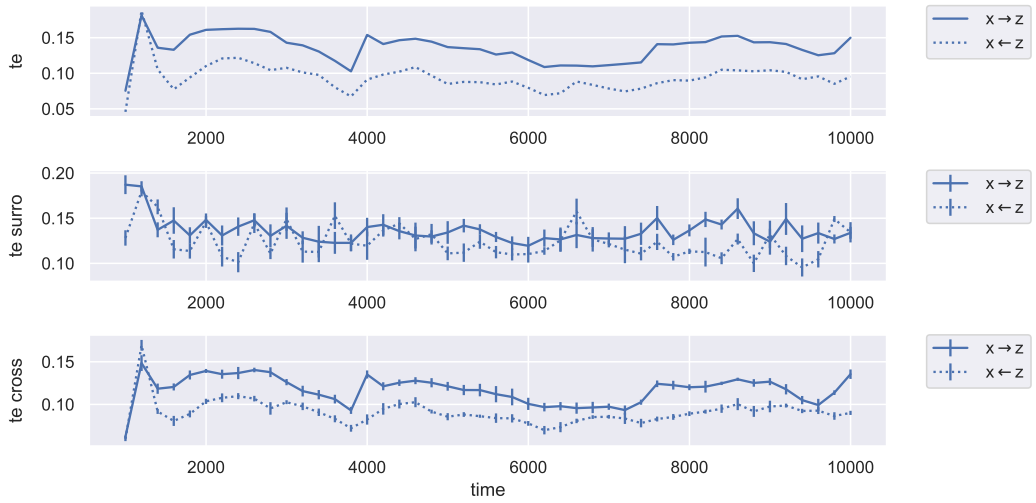

Figure 5.47: TE between x and z for short timeframes in the Lorenz system.

Convergent Cross Mapping

While for CCM we also observe similar evolutions for both directions of the x and y pair, we do not see similarity to the cross evolutions. As depicted in Figure 5.48, the causality for both original and surrogate data is generally high with fluctuations and one significant outlier for the original data. This is caused by the low density of the attractor for which CCM fails in some cases.

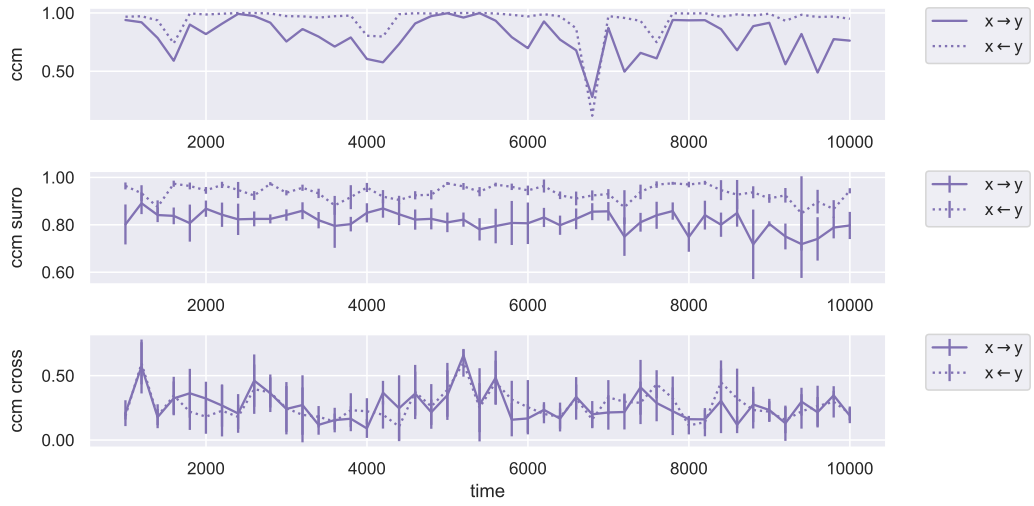

Figure 5.48: CCM between x and y for short timeframes in the Lorenz system.

Heavy fluctuations can also be observed for the original data of the y and z pair in Figure 5.49. However, the surrogate evolutions are generally vanishing with two distinct outliers which can be attributed to periodic structures in these phases.

Similarly to TE, Figure 5.50 again confirms the similarity between the evolutions of z and the other coordinates.

In conclusion, CCM shows a rather consistent causality evolution structure for both short and large timeframes. Despite of the higher fluctuations, we see that the causality for z and the other coordinates is driven by nonlinearity, while the x and y pair remains unaffected by surrogatization.

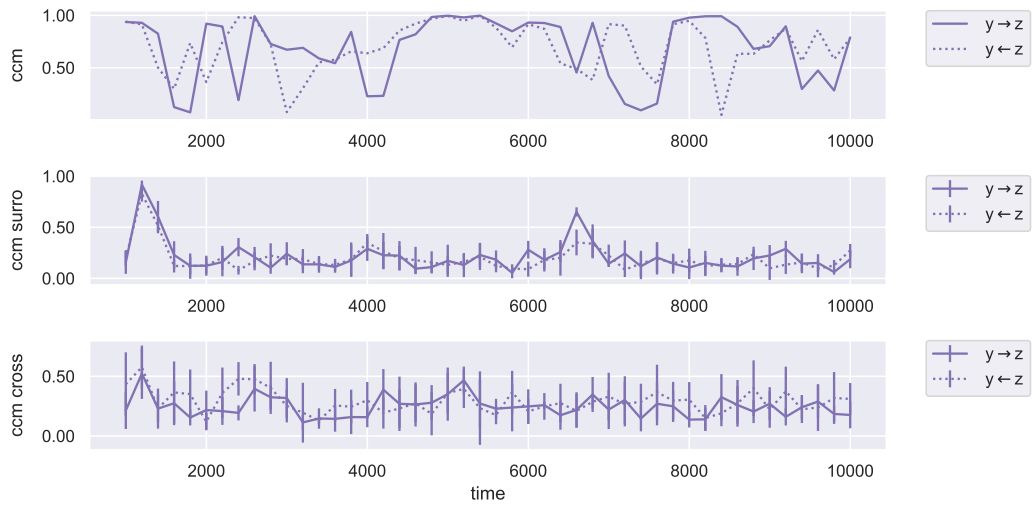

Figure 5.49: CCM between y and z for short timeframes in the Lorenz system.

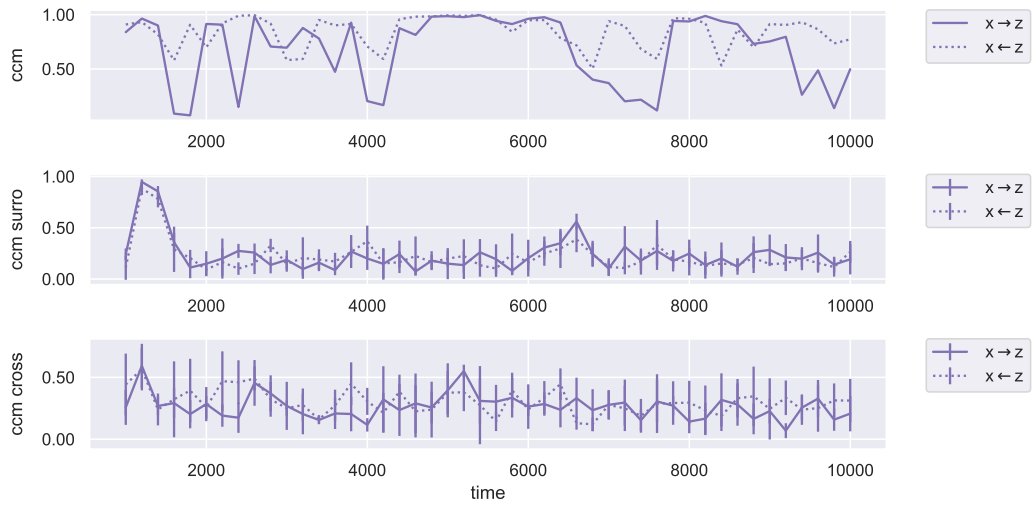

Figure 5.50: CCM between x and z for short timeframes in the Lorenz system.

Mapping

Due to the better resolution provided by the short timeframes, we are able to map the causality evolutions to the attractor. Therefore, we rescale the values to a range between $(0, 1)$ in order to have the same color gradient. While we use the system causality for the map, the bivariate evolutions could also be taken. We leave this to further research.

The system evolutions are depicted in Figure 5.51. As indicated in the previous sections, the original and cross TE have almost identical paths confirmed by a nested correlation of $\rho = 0.95$. In contrast, CCM is generally higher for original than for surrogate data.

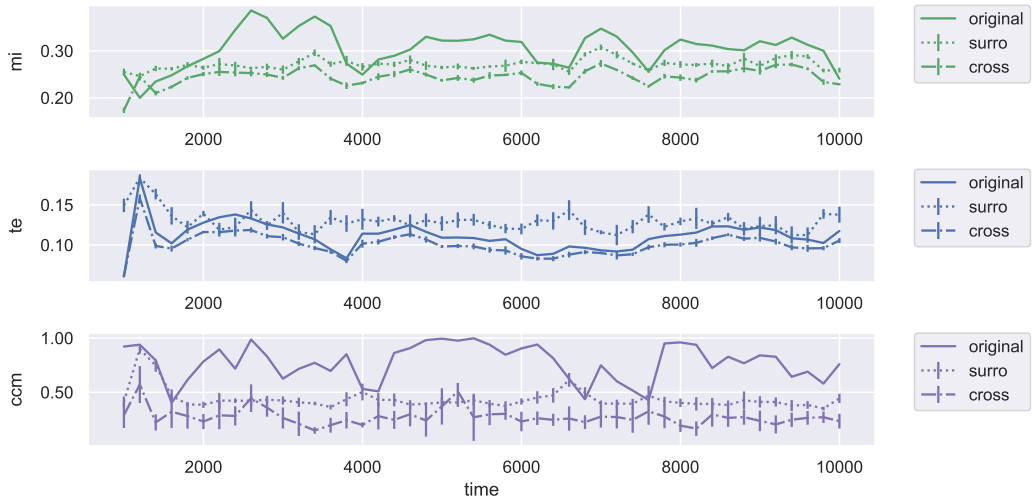

Figure 5.51: Absolute norm for short timeframes in the Lorenz system.

Figure 5.52 depicts the absolute norm of TE mapped to the Lorenz attractor, for which we see a rather even pattern with medium causality except for the inner circle of the left wing. This area seems to have alternating circles of high and low TE.

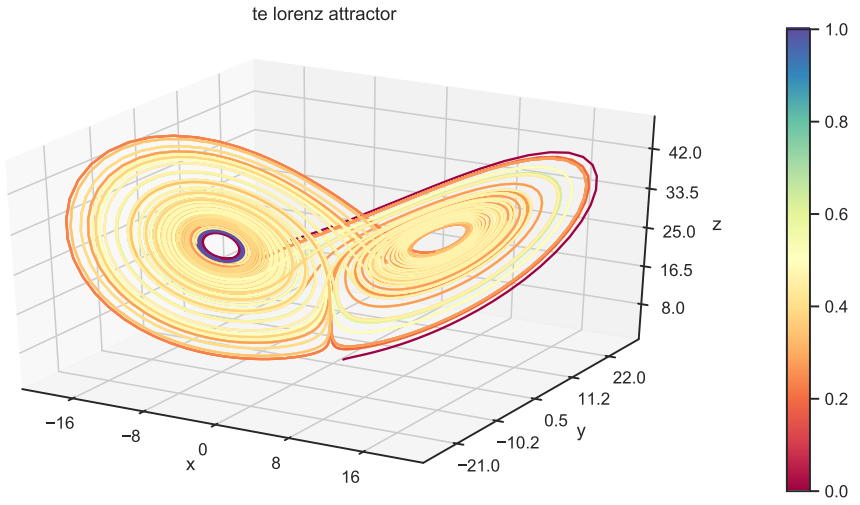

Figure 5.52: Absolute norm TE mapped to the Lorenz attractor.

The corresponding surrogate map is illustrated in Figure 5.53, where the left wing exhibits low causality except for the inner circle. The right wing seems to generally have a higher level of causality than the left one.

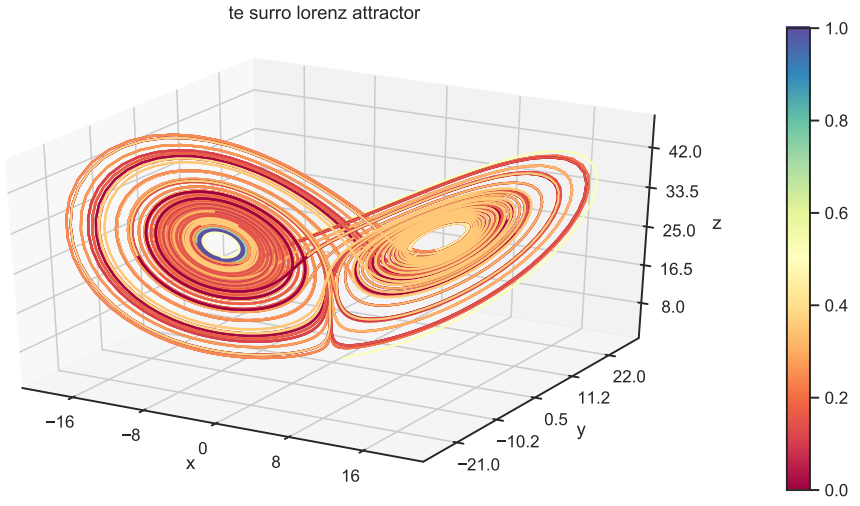

Figure 5.53: Absolute norm surrogate TE mapped to the Lorenz attractor.

The absolute norm CCM map is illustrated in Figure 5.54, where the right wing shows an even pattern of medium causality. In contrast, the left wing has alternating circles of high and medium causality and again shows larger fluctuations between high and low causality in the inner circles.

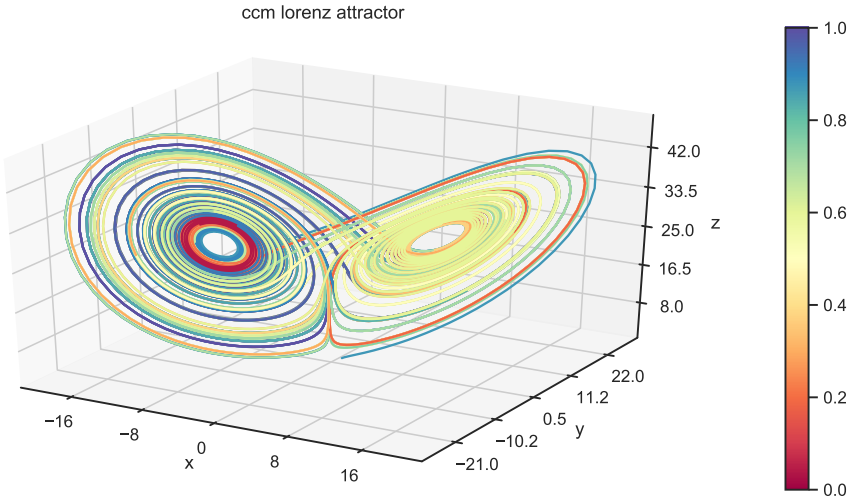

Figure 5.54: Absolute norm CCM mapped to the Lorenz attractor.

The CCM causality structure for surrogate data appears to be uniformly low with an exception in the inner center of the left wing of the attractor. Again, we observe alternations between low and high causality.

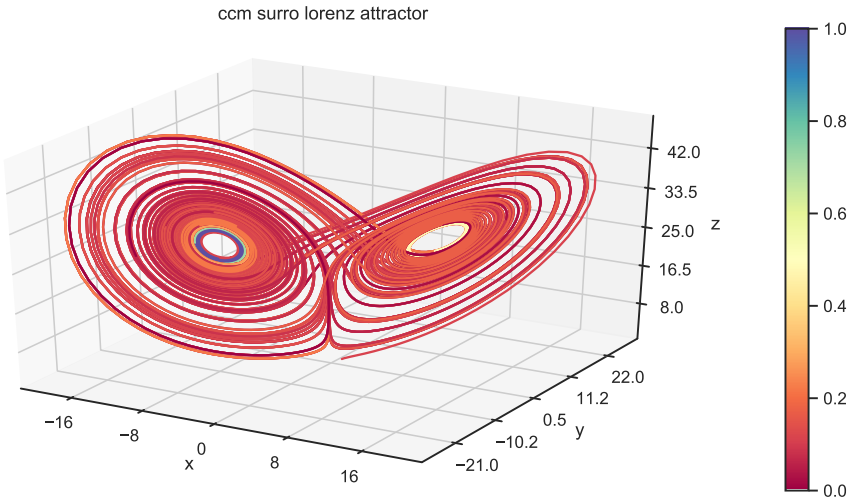

Figure 5.55: Absolute norm surrogate CCM mapped to the Lorenz attractor.

Chapter 6

Conclusion

This thesis presents a novel framework for detecting causality drivers in time series systems based on the three inference techniques: *Granger Causality*, *Transfer Entropy*, and *Convergent Cross Mapping*.

Methodically, it introduces a continuous scoring scale for Granger Causality which replaces the hypothesis tests frequently used in the literature. Furthermore, a normalization for Transfer Entropy is suggested. Regarding Convergent Cross Mapping, a convergence dependent scoring method which automates the causality inference s developed and makes this technique applicable to large scales.

By using Fourier Transform surrogates this model enables the detection of linear and nonlinear causality drivers. Therefore, it introduces cross measures, which in combination with original and surrogate measures can help to attribute the causality to the linear or nonlinear properties of the individual time series. Furthermore, several nonlinearity measures are defined, which quantify different aspects and interpretations of nonlinearity.

In order to evaluate higher dimensional time series systems, new measures are created based on widely used matrix norms. These concepts are extended to measure asymmetry in systems. For causal systems a new measure is presented, which is inspired by Markov chains and incorporates indirect causal couplings.

Additionally, this thesis introduces nested measures, which are calculated on the causality evolutions and enable confirming or finding new relations between statistical and causality measures.

The empirical analysis performed on synthetic nonlinear systems confirms that Granger Causality, Transfer Entropy, and Convergent Cross Mapping measure different aspects of causality and hence need to be evaluated together in order to make valid statements on causality drivers. While Granger Causality serves as a backtest for linear causality, Transfer Entropy and Convergent Cross Mapping offer different

benefits. A structured parameter grid analysis shows that Transfer Entropy detects coupling strengths more reliably, whilst Convergent Cross Mapping is more robust to noise and offers a convenient normalization.

The analysis on the coupled logistic system and Lorenz attractor indicates that causality is mainly driven by nonlinear properties for larger timeframes. It further suggests that the original, surrogate, and cross causality at system level remain constant with the original causality being significantly higher than the other two. Thus, a general conclusion of this thesis is that the causality on system level remains constant at large timeframes and is mainly driven by nonlinear properties. While this statement is confirmed for some other synthetic nonlinear systems, further research needs to be conducted.

In order to identify the causality structure within the Lorenz attractor, an analysis is performed with timeframes which are purposely chosen too short. However, this allows mapping causality strengths to locations within the attractor. It is shown that the inner circle of the left wing has an alternating causality structure. Though, in order to make generally valid statements, this approach needs to be subject to further research.

Regarding information theoretic measures, this thesis indicates that the probability density estimation using histograms depends strongly on the binning resolution within the distribution core ranges. However, it encourages further research on a universal rule in order to ensure robust estimations. Until this is achieved, results deduced from Transfer Entropy must be considered with caution.

Since the model developed as part of this thesis is designed to be applicable to real-world time series systems, it has the potential to generate new findings across all sciences.

Bibliography

[1] George Sugihara et al. "Detecting causality in complex ecosystems". In: *science* 338.6106 (2012), pp. 496–500.

[2] Alexander Haluszczynski et al. "Linear and nonlinear market correlations: Characterizing financial crises and portfolio optimization". In: *Physical Review E* 96.6 (2017), p. 062315.

[3] Christoph Räth and R Monetti. "Surrogates with random Fourier phases". In: *Topics on Chaotic Systems: Selected Papers from Chaos 2008 International Conference*. World Scientific. 2009, pp. 274–285.

[4] Thomas Schreiber and Andreas Schmitz. *Improved surrogate data for nonlinearity tests*. 1999. arXiv: `chao-dyn/9909041 [chao-dyn]`.

[5] Dean Prichard and James Theiler. "Generating surrogate data for time series with several simultaneously measured variables". In: *Physical review letters* 73.7 (1994), p. 951.

[6] Jakob Runge. "Causal network reconstruction from time series: From theoretical assumptions to practical estimation". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.7 (2018), p. 075310.

[7] Thomas Schreiber. "Measuring information transfer". In: *Physical review letters* 85.2 (2000), p. 461.

[8] Lionel Barnett, Adam B Barrett, and Anil K Seth. "Granger causality and transfer entropy are equivalent for Gaussian variables". In: *Physical review letters* 103.23 (2009), p. 238701.

[9] Anastasios A Tsonis et al. "Convergent cross mapping: theory and an example". In: *Advances in Nonlinear Geosciences*. Springer, 2018, pp. 587–600.

[10] J Huke. "Embedding nonlinear dynamical systems: A guide to takens' theorem (Technical Report)". In: *Manchester Institute for Mathematical Sciences, University of Manchester* (1993).

[11] Matthew B Kennel, Reggie Brown, and Henry DI Abarbanel. "Determining embedding dimension for phase-space reconstruction using a geometrical construction". In: *Physical review A* 45.6 (1992), p. 3403.

[12] *rEDM: An R package for Empirical Dynamic Modeling and Convergent Cross Mapping*. `https://mran.revolutionanalytics.com/snapshot/2018-07-06/web/packages/rEDM`. Accessed: 2020-08-20.

Erklärung zur Masterarbeit

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

Master's thesis statement of originality

I hereby confirm that I have written the accompanying thesis by myself, without contributions from any sources other than those cited in the text and acknowledgments.

.....	.....
Date and location	Signature