Reconstructing Dynamical Systems From Stochastic Differential Equations to Machine Learning

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)
im Fach Physik
Spezialisierung: Theoretische Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
MSc. Forough Hassanibesheli,

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr. Julia von Blumenthal

Dekanin der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Caren Tischendorf

Gutachter:

1. Prof. Dr. Dr. h.c. mult. Jürgen Kurths
2. Prof. Dr. Milan Palus
3. Prof. Dr. Peter Ditlevsen

Tag der mündlichen Prüfung: 27.01.2023

Abstract

Modeling complex systems with large numbers of degrees of freedom have become a grand challenge over the past decades. Typically, only a few variables of complex systems are observed in terms of measured time series, while the majority of them – which potentially interact with the observed ones - remain hidden. Hence, to analyze and model the behavior of such systems from available data, a broad spectrum of data-driven methods ranging from network theory to machine learning have been proposed. Throughout this thesis, we tackle the problem of reconstructing and predicting the underlying dynamics of complex systems using different data-driven approaches.

In the first part, we address the inverse problem of inferring an unknown network structure of complex systems, reflecting spreading phenomena, from observed event series. For this purpose, we numerically investigate two types of event-based processes; I) a general model of events propagation with spontaneous events and triggered events, and II) Susceptible-Infected-Recovered-Susceptible (SIRS) model of epidemic spreading. Then we study the pairwise statistical similarity between the sequences of event timings at all nodes through event synchronization (ES) and event coincidence analysis (ECA), relying on the idea that functional connectivity can serve as a proxy for structural connectivity. We demonstrate that both ES and ECA indeed can accurately infer the underlying network structure from the timing of events without using any prior knowledge of the type of observed spreading dynamics.

In the second part, we focus on reconstructing the underlying dynamics of complex systems from their dominant macroscopic variables using different Stochastic Differential Equations (SDEs). Since the microscopic dynamics of complex systems are often not accessible, SDEs attempt to model the macroscopic variables explicitly and represent the microscopic variables in terms of noise. We investigate the performance of three different SDEs – the Langevin Equation (LE), Generalized Langevin Equation (GLE), and the Empirical Model Reduction (EMR) approach in this thesis. Our results reveal that LE demonstrates better results for systems with weak memory while it fails to reconstruct underlying dynamics of systems with memory effects and colored-noise forcing. In these situations, the GLE and EMR are more suitable candidates since the interactions between observed and unobserved variables are considered in terms of memory effects. Later, we conduct a data-driven analysis of the Greenland temperature and atmospheric circulation proxies under the purview of stochastic processes. Our results unravel the features of the climate system's stability landscape that helps understand candidate mechanisms underlying these abrupt climate changes.

After gaining knowledge about the underlying dynamics of complex systems, the next step is to predict their behavior. In the last part of this thesis, we develop a model based on the Echo State Network (ESN), combined with the past noise forecasting (PNF) method, to predict real-world complex systems. One of the highly complex physical systems is the Earth's climate system which consists of several interconnected subsystems. We attempt to forecast the behavior of various climate oscillations such as the El Nino/Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), and the Atlantic Multidecadal Oscillation (AMO) using our ESN-based model. Our results show that the proposed model captures the crucial features of the underlying dynamics of climate variability. Moreover, the predictive power of our model suggests that cross-scale interactions are indeed important for accurately modeling these systems.

Zusammenfassung

Die Modellierung komplexer Systeme mit einer großen Anzahl von Freiheitsgraden ist in den letzten Jahrzehnten zu einer großen Herausforderung geworden. In der Regel werden nur einige wenige Variablen komplexer Systeme in Form von gemessenen Zeitreihen beobachtet, während die meisten von ihnen - die möglicherweise mit den beobachteten Variablen interagieren - verborgen bleiben. Um das Verhalten solcher Systeme anhand der verfügbaren Daten zu analysieren und zu modellieren, wurde daher ein breites Spektrum datengesteuerter Methoden vorgeschlagen, die von der Netzwerktheorie bis zum maschinellen Lernen reichen. In dieser Arbeit befassen wir uns mit dem Problem der Rekonstruktion und Vorhersage der zugrunde liegenden Dynamik komplexer Systeme mit Hilfe verschiedener datengestützter Ansätze.

Im ersten Teil befassen wir uns mit dem inversen Problem, eine unbekannte Netzwerkstruktur eines komplexen Systems, das Ausbreitungsphänomene widerspiegelt, aus beobachteten Ereignisreihen abzuleiten. Zu diesem Zweck untersuchen wir numerisch zwei Arten von ereignisbasierten Prozessen: I) ein allgemeines Modell der Ereignisausbreitung mit spontanen und ausgelösten Ereignissen und II) das SIRS-Modell (Susceptible-Infected-Recovered-Susceptible) der epidemischen Ausbreitung. Anschließend untersuchen wir die paarweise statistische Ähnlichkeit zwischen den Sequenzen der Ereigniszeitpunkte an allen Knoten durch Ereignissynchronisation (ES) und Ereignis-Koinzidenz-Analyse (ECA), wobei wir uns auf die Idee stützen, dass funktionale Konnektivität als Stellvertreter für strukturelle Konnektivität dienen kann. Wir zeigen, dass sowohl ES als auch ECA in der Tat die zugrunde liegende Netzwerkstruktur aus dem Timing von Ereignissen ableiten können, ohne vorheriges Wissen über die Art der beobachteten Ausbreitungsdynamik zu verwenden.

Im zweiten Teil der Arbeit konzentrieren wir uns auf die Rekonstruktion der zugrundeliegenden Dynamik komplexer Systeme aus ihren dominanten makroskopischen Variablen unter Verwendung verschiedener stochastischer Differentialgleichungen (SDEs). Da die mikroskopische Dynamik komplexer Systeme oft nicht zugänglich ist, versuchen SDEs, die makroskopischen Variablen explizit zu modellieren und die mikroskopischen Variablen als Rauschen darzustellen. In dieser Arbeit untersuchen wir die Leistung von drei verschiedenen SDEs - der Langevin-Gleichung (LE), der verallgemeinerten Langevin-Gleichung (GLE) und dem Ansatz der empirischen Modellreduktion (EMR). Unsere Ergebnisse zeigen, dass die LE bessere Ergebnisse für Systeme mit schwachem Gedächtnis zeigt, während sie die zugrunde liegende Dynamik von Systemen mit Gedächtniseffekten und farbigem Rauschen nicht rekonstruieren kann. In diesen Situationen sind GLE und EMR besser geeignet, da die Wechselwirkungen zwischen beobachteten und unbeobachteten Variablen in Form von Speichereffekten berücksichtigt werden. Später führen wir eine datengestützte Analyse der grönländischen Temperatur- und atmosphärischen Zirkulationsproxies unter dem Gesichtspunkt stochastischer Prozesse durch. Unsere Ergebnisse entschlüsseln die Merkmale der Stabilitätslandschaft des Klimasystems, was zum Verständnis der möglichen Mechanismen beiträgt, die diesen abrupten Klimaänderungen zugrunde liegen.

Nach dem Erwerb von Kenntnissen über die zugrunde liegende Dynamik komplexer Systeme besteht der nächste Schritt darin, ihr Verhalten vorherzusagen. Im letzten Teil dieser Arbeit entwickeln wir ein Modell auf der Grundlage des Echo State Network (ESN) in Kombination mit der PNF-Methode (Past Noise Forecasting), um komplexe Systeme in der realen Welt vorherzusagen. Eines der hochkomplexen physikalischen Systeme ist das Klimasystem der Erde, das aus mehreren miteinander verbundenen Teilsystemen besteht. Wir versuchen, das Verhalten verschiedener Klimaschwingungen wie der El Nino/Southern Oscillation (ENSO), der Pacific Decadal Oscillation (PDO) und der Atlantic Multidecadal Oscillation (AMO) mit unserem ESN-basierten Modell vorherzusagen. Unsere Ergebnisse zeigen, dass das vorgeschlagene Modell die entscheidenden Merkmale der

zugrunde liegenden Dynamik der Klimavariabilität erfasst. Darüber hinaus deutet die Vorhersagekraft unseres Modells darauf hin, dass skalenübergreifende Wechselwirkungen tatsächlich wichtig für die genaue Modellierung dieser Systeme sind.

List of Publications

This dissertation is partly based on the following publications.

I) **Hassanibesheli, F.**, Donner, R. V. (2019). Network inference from the timing of events in coupled dynamical systems. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(8), 083125.

II) **Hassanibesheli, F.**, Boers, N., and Kurths, J. (2020). Reconstructing complex system dynamics from time series: a method comparison. New Journal of Physics, 22(7), 073053.

III) Rydin, L., Riechers, K., **Hassanibesheli, F.**, Witthaut, D., Lind, P. G., and Boers, N. (2021). Data-driven Reconstruction of Last Glacials' Climate Dynamics Suggests Monostable Greenland Temperatures and a Bistable Northern Hemisphere Atmosphere. Earth Syst. Dynam, in review

IV) **Hassanibesheli, F.**, Boers, N., and Kurths, J. (2021). Long-term ENSO Prediction with Echo-State Network. Environmental Research: Climate, in review

V) **Hassanibesheli, F.**, Boers, N., and Kurths, J. (2022). Predicting Climate Oscillations Using Echo-State Network, in preparation

Acknowledgements

Now that this journey is nearing its end, I am delighted to be able to look back and remember all the people who helped me; among all of them, I would like to thank some of them in particular.

First and foremost, I would like to give my deepest gratitude to my supervisors, Professor Jürgen Kurths and Professor Niklas Boers, for guiding me throughout my Ph.D., from concept to completion. This journey was not possible without their insightful discussions and constant intellectual support.

I would also like to thank Max and Frederik for their time and valuable comments that helped me to improve this thesis. I would like to express my sincere thanks to Professor Reik Donner, who I collaborated with on the first project presented in this thesis. And Thanks to the Deutscher Akademischer Austauschdienst (DAAD) for providing financial and administrative support. support.

I owe a lot to my family in every aspect of my life. I would like to extend my special thanks to my beloved Dad and Mom, Faramarz and Shahrbanoo, and my dearest sisters, Parvin and Negar, for making me believe in myself and giving me the strength to pursue my dreams.

Last but not least, I want to express my heartfelt gratitude to my spouse, Anoushiravan, for his endless love and innumerable sacrifices. I am incredibly lucky to have him all by my side. I could not have done it without him!

Contents

List of Figures

List of Tables

List of frequently used abbreviations

SDE	Stochastic Differential Equation
LE	Langevin Equation
KM	SKramers–Moyal Expansion
GLE	Generalized Langevin Equation
MZ	Mori-Zwanzig Formalism
EMR	Empirical Model Reduction
ANN	Artificial Neural Network
ESN	Echo State Networks
PNF	Past Noise Forecasting
ES	Event Synchronization
ECA	Event Coincidence Analysis
PDF	Probability Density Functions
AFC	Auto-correlation Function
PCC	Pearson's Correlation Coefficient
RMSE	Root Mean Squared Error
POD	Probability of Detection
HSS	Heidke-Skill Score
ROC	Receiver Operating Characteristic
TPR	True Positive Rate
FPR	False Positive Rate
OU	Ornstein–Uhlenbeck
ENSO	El Niño-Southern Oscillation
PDO	Pacific Decadal Oscillation Prediction

1. Introduction

" How do we know that the world's creation is not determined by falling grains of sand" (Victor Hugo 1862). Consider a single grain of sand in a sand pile; any change in its position can cause a massive avalanche, altering the sand pile's shape. Systems with many coupled degrees of freedom, whose macroscopic dynamics cannot easily be predicted from the behavior of the individual components, are called complex systems [152]. Ecosystems, financial markets, granular materials, and the Internet [11, 192, 204] are all well-known instances of complex systems. The prevalence of these systems and their colossal influences on human lives, highlight the theoretical and transnational importance of understanding, modeling, and predicting them.

The complexity of complex systems does not come only from many interacting entities but also from nonlinear interactions that constantly change their internal patterns and structures. Unlike simple linear systems, where an ensemble of individual features can provide complete knowledge of the overall dynamics, in complex systems, the whole is more than the sum of the components. Notably, complex systems exhibit some distinct features that make their dynamics challenging to investigate, such as adaptiveness, unexpected or unpredictable emergence, sensitivity to initial conditions, and self-organization [37, 187]. For instance, short- to medium-range weather forecasting is prone to error due to sensitivity to initial conditions [130]. Due to this intrinsic complexity and interactions in a wide range of time scales, capturing detailed local knowledge of every microscopic degree of freedom is difficult. However, in the study of complex systems, we are often interested in global emergent behaviors that occur at a larger scale than the microscopic variables' behavior. To derive such macroscopic-level information, one possible approach is to observe the system at a particular time scale and abstract away unnecessary details on the finer scale that cannot be resolved explicitly [192]. For example, to capture the essential information on the global climate system, one can investigate the evolution of sea surface temperature (SST) instead of understanding each microscopic component of oceans and atmosphere.

Describing the long-term dynamics of a complex system by macroscopic variables fluctuating in time is the heart of stochastic modelling [75, 138]. Depending on the time scale differences between microscopic and macroscopic variables, one can achieve a stochastic description of a system's dynamic based on a desired level of abstraction. In scenarios where the macroscopic behavior of a system occurs at a far larger scale than the microscopic variables, one can study macroscopic variables independently and consider microscopic variations as noise (i.e., Markov processes). For instance, it has been demonstrated that one can consider turbulent free jet as a Markov process and establish a model based on the Fokker-Planck equation [57]. However, this scale separation approach cannot be easily applied when the macroscopic behavior depends on the nonlinear microscopic interactions, which can lead to long memory effects [123]. For example, one can consider the Earth's climate system which consists of several interconnected subsystems (i.e., oceans and atmosphere) taking place on different time and spatial scales. Even though these subsystems may be investigated separately, our growing understanding of nonlinear interactions and feedback loops among them demonstrates that such interrelations must be taken into account to understand the emergent behaviors of the Earth's system [46]. Accordingly, to capture these cross-scale interactions, a model needs to have additional terms that consider

memory effects. For example, the Mori–Zwanzig [154, 225] formalism has been developed to construct a coarse-grained model from high-dimensional systems and derive closed equations of motion of resolved variables by considering cross-scale interactions. However, analyzing the dynamics of a complex system using such models can be a complicated and no-trivial task. Additionally, establishing mathematical models can be challenging for complex dynamical systems where systems' dynamics are driven by non-Gaussian processes (e.g., Lévy motion).

As an alternative approach, data-driven techniques can be employed to extract hidden patterns of an unknown complex system from the evolution of observed data. Different data-driven methods, from network science to time series analysis and machine learning, have been introduced to identify emergent properties of complex systems [27, 71]. For example, in climate science, data-driven methods, successfully are employed for weather forecast [29, 178] and climate variability prediction [72, 149, 219]. A major advantage of these data-driven approaches is that they do not require a prior assumption about the dynamics of the target system. The overarching aim of the current thesis is to shed light on data-driven approaches to understand and reconstruct the underlying dynamics of complex systems and subsequently predict their future behavior. More specifically, the current thesis will address three questions:

I. Is it possible to reconstruct the structural connectivity of an unknown complex network from observed event-series?

One common approach to represent an abstraction of a system is a network. Complex networks are widely applied for modeling and analyzing real-world phenomena across disciplines [51]. In real-world situations, the network structure underlying some observed macroscopic dynamics is often unknown. In such cases, the common task is to infer the unobserved connectivity patterns based on long-term observations of the emerging dynamics. The fundamental assumption underlying this concept is that strong pairwise statistical associations between the dynamics at each pair of nodes can be interpreted as functional connectivity between them [78, 79, 208]. This functional connectivity can be used to proxy the (unobservable) actual structural connectivity of the network. However, the associated problem of statistical inference of the underlying linkage structure – the so-called network topology – from observed dynamical processes is still a subject of ongoing research. Specifically, only a few studies have deeply addressed whether we could correctly "predict" the placement of connections from specific measures of statistical similarities between event sequences [44, 217]. To solve the problem of inferring structural connectivity of the target systems, we will use suitable similarity measures, i.e., event synchronization and event coincidence analysis.

II. How to reconstruct the underlying dynamics of complex systems form time series using stochastic differential equation (SDE)?

Due to insufficient prior knowledge about the underlying physical processes, deriving governing laws (in the forms of ordinary, partial, or stochastic differential equations) often analytically is difficult [139]. In the last several decades, there have been considerable methodological advances in discovering governing laws underlying the evolution of complex systems from experimental and observational data [23, 25, 104]. The time evolution of a complex system's state can be a continuous trajectory, exhibit jump discontinuities, or even in some cases, stochastic. Notably, the stochasticity might stem from random forcing or the presence of interactions in different time (length) scales. This is where the concept of stochastic models becomes more relevant to describing the underlying dynamics of a system. Stochastic modeling is commonly em-

ployed in domains such as finance [49, 52], power grid [8, 65] and climate systems [19, 137]. In this framework, the evolution of a complex system can be described by its macroscopic dynamics. Further, the microscopic interactions are accounted as fluctuating forces modeled by suitable noise terms. The noise term serves as a tool for taking non-specified components into account that cannot be captured by a deterministic model [181]. Notably, depending on the intrinsic dynamics of the target system (e.g., from weak history-dependence to strong state-dependence), different forms of the stochastic equation of motion can be derived. We will employ some data-driven methods to extract underlying mechanism and corresponding SDEs of complex systems.

III. How to improve the prediction of future behavior of complex systems from limited time series using artificial neural networks (ANNs)?

Aside from identifying, predicting complex systems' behavior is essential in many fields, particularly in Earth's climate system (e.g, forecasting weather and extreme events). However, due to the complex nature of such systems and often their chaotic characteristics (i.e., high sensitivity to initial conditions), there is an intrinsic limit to predicting their behavior. In recent years, different machine learning techniques, particularly artificial neural networks (ANNs), have been advanced as popular tools to classify, optimize, and diagnose the characteristics of complex systems [129, 212]. The central issue of these techniques is how to extract useful information efficiently from available data. However, a pivotal problem in forecasting natural systems, like climate, is understanding the influence of unresolved (high-frequency variables) on the dynamics of resolved (low-frequency variables) [10]. To circumvent this problem, we developed an ANN algorithm that is not only able to predict the variable of interest but can also estimate fast fluctuations and their impact on the resolved parts of the dynamics.

1.1. Organization of the thesis

This thesis is split into two parts: the theory and applications. In chapter (2), I start with an overview of the basic concepts of complex networks and explain how functional connectivity can be estimated based on the temporal similarity measures of event sequences. In the second section, I introduce several dimension-reduction methods commonly used to reconstruct the governing dynamics of stochastic systems. Finally, I close the theoretical foundation chapter with an overview of ANNs and present our developed ANN based on the interactions between the slow and the fast scales in real-world data. From chapter (3) onward, I present our original results employing these theoretical foundations for addressing above-mentioned questions. In chapter (3), we aim to infer the (unknown) structural connectivity of complex networks, exhibiting some spreading process, from the functional connectivity. In chapter (4), first, we reconstruct SDEs governing the dynamics of various synthetics and real-world processes. Then, we investigate the underlying stochastic process deriving sudden climatic transitions observed during the last glacial period. In chapter (5), we evaluate the prediction skill of the developed ANN by applying it to three different climate variability indices. Finally, in chapter (6), I summarize the main conclusions derived from chapter (3) to chapter (5) and outline potential avenues for future research.

Part I.

Theoretical Framework

2. Theoretical Foundations

2.1. Measures of Similarity in Complex Networks

We live in a world of complex systems. It is usually challenging to model such systems due to their highly complex nature and many interdependencies. One popular approach used to extract information from structural and behavioral characteristics of complex systems is complex network theory [10, 22, 163, 196]. According to this theory, many complex systems ranging from biological and social systems to ecosystems and the Internet can be described as networks of interacting entities. This powerful approach that has traditionally been the domain of graph theory, attempts to uncover the signature behavior of complex systems using a visual-mathematical abstraction of a set of interactions and removes all unnecessary details. However, constructing the structural network of many real-world systems (e.g., climate and brain) is not always straightforward. In such circumstances, the unknown connectivity patterns should be inferred from the statistical interrelationships among entities of a network [81]. The fundamental assumption underlying this concept is that strong pairwise statistical associations between two entities can be interpreted as functional connectivity between them, which can be used as a proxy of the (unobservable) actual structural connectivity [26, 45, 47, 164]. To capture such functional connectivity, different similarity measures have been proposed [44, 174]. In the following section, we briefly introduce some fundamentals of network theory and present some similarity measures used in this thesis to infer network topology.

2.1.1. Complex Networks

Mathematically, a network consists of a set of nodes $\mathcal{N}$ and links $\mathcal{E}$ that represent various entities or systems. Properties of a network can be characterized by its connectivity matrix– known as adjacency matrix $\mathcal{A}$ – which keeps tracking of the topological structure of the network [16].

$$A_{ij} = \begin{cases} 1 & \text{if there is a link from } i \text{ to } j \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

In an undirected network, the adjacency matrix is symmetric, $A_{ij} = A_{ji}$ indicating a connection between two nodes is bidirectional. However, in applications, complex networks are not always symmetric, and a link between two nodes has a specified orientation $A_{ij} \neq A_{ji}$ (e.g., neural connections in the brain and internet). For such directed networks, a link from a source node i to a target node j is not equivalent to a connection from j to i. Moreover, a network can have self-loops, where the $A_{ii} = 1$. In both directed and undirected networks, the strength of links between nodes can also be represented by weights (W_{ji}). Such weighted physical linkages are the basis of the architecture of weighted networks where the entries of the adjacency matrix are: $A_{ij} = W_{ij}$. In addition, each node in a network can possess a weight that represents how significantly a node contributes to a domain of interest.

Several global and local measures can be defined as indicators of the network's topology using the adjacency matrix. For instance, we can quantify the number of edges on a network using the link density ρ.

$$\rho = \frac{\sum_{ij} a_{i,j}}{N(N-1)} \tag{2.2}$$

Here the denominator represents the maximum number of edges that could potentially exist in a given network. Note that, for an undirected network, the numerator must be multiplied by 2.

A network's average shortest path length is an effective method for learning how a network is arranged and connected. This quantity measures the average number of steps with the shortest paths for all possible pairs of network nodes.

$$< L(i,j) >= \frac{1}{N(N-1)} \sum_{i \neq j}^{N} L_{i,j} \tag{2.3}$$

Another frequently used tool for analyzing a network's topology is the clustering coefficient, representing the probability that two randomly selected neighbors of a node are connected. This metric, known as local clustering coefficients, is defined as follows:

$$< C(i,j) >= \frac{n_i}{k_i(k_i-1)} \tag{2.4}$$

Where k_i denotes the degree of node i and n_i stands for the number of links that connect the neighbors of node i. We can calculate the global clustering coefficient by averaging this measure for all nodes in the network.

The question arising here is how to interpret these measurements for an empirical network in a meaningful manner. Random network models provide suitable reference points for determining whether the measured properties of a network of interest are exceptional. This framework treats links between nodes as random variables governed by specific probabilistic laws. Three types of random network models are considered in this thesis: Erdös-Rényi random graphs, scale-free Barabási-Albert networks [12](SF), and small-world Watts-Strogatz networks [214] (SW). The Erdös-Rényi model (ER) generates fully random networks in which a possible edge between two pairs of nodes is chosen randomly with a certain probability (p). Random networks evolved under the ER model exhibit small average shortest path length and small clustering coefficient. On the other hand, the small-world network (SW) is a class of random graphs with a rewiring probability ρ of the original underlying ring lattice. In this structure, the edge between two nodes is disconnected and then randomly connected to another node (see Fig. 2.1). By varying the ρ, one can study the transition of the network from a ring lattice to a random structure.

In contrast to the ER and SW networks, a SF network's degree distribution (P) exhibits a power-law behavior where P decays as the degree k increases. The algorithm of generating ER and SW networks is based on a network with a fixed number of nodes, while the common models to construct a SF network starts with an initial small network with n nodes and k degree. Then, at each time step t, a new node is added into the network and connected to m existing nodes chosen with a probability proportional to their current degree.

2.1.2. Similarity Measures

Determining pairwise similarity is a key problem when dealing with time series. Such inter-dependencies can be characterized by many different measures, the use of which

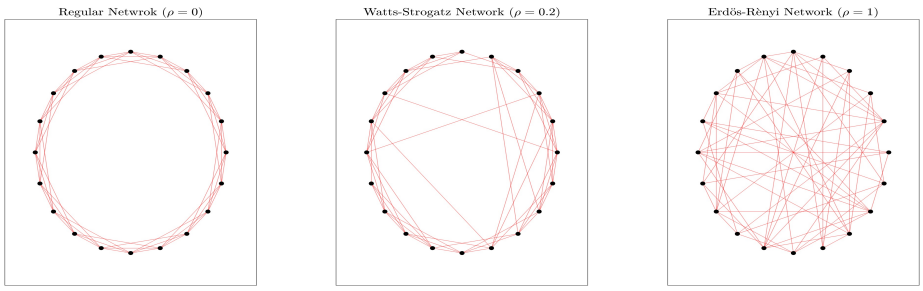

FIGURE 2.1.: Representation of Watts-Strogatz network as an intermediary state between regular and random network, from left to right.

should be guided by the specific research question and type of recorded observables. For example, Pearson's correlation coefficient (PCC) [45] is one of the most commonly used similarity metrics. The PCC is restricted to measuring linear dependencies' strength between two normalized time series. It can take values between -1 and 1, where the positive and negative signs correspond to the direction of the linear relationship. For two time series x^r and x^r of length T, the PCC can be defined as follows:

$$PCC(x^r, x^p) = \frac{\sum_{i=1}^{n}(x_i^r - \overline{x}_i^r).(x_i^p - \overline{x}_i^p)}{\sigma_{x^r}.\sigma_{x^p}} \tag{2.5}$$

where σ_{x^r} and σ_{x^p} are the standard deviation. The PCC(x^r, x^p) = 0 indicates that the two time series are perpendicular and are not correlated.

In many complex systems, the pertinent information is carried by specific events in a time series, for instance, heavy rainfall events in daily precipitation records or neuronal spikes in electroencephalogram (EEG) signals. In such situations where the variables of interest are not continuous, the classical linear statistical association measures may not provide reliable information on the co-occurrence of events. Among the recent methodological developments, Event synchronization (ES) and Event coincidence analysis (ECA) are known as non-linear similarity measures suitable for binary data. These methods allow to identify instantaneous and delayed coupling patterns among paired event time series.

2.1.2.1. Event Synchronization (ES)

Event synchronization (ES) was initially introduced by Quian Quiroga [174] as a parameter-free method to measure the strength of mutual synchronization (and associated time delay) between neurophysiological signals exhibiting spiky dynamics (like in electroencephalogram (EEG) recordings). Beyond various applications in the neuroscience context, this approach has also been applied to climatological time series in recent years. For instance, to investigate the synchronicity of the timing of extreme events at different locations, to identify essential regional interdependence patterns and the preferred direction of propagation of such extremes [17, 18, 139].

Let us consider two series containing information on the times of occurrence of two specific types of events, λ and μ (e.g., heavy rainfall events at different meteorological measurement stations). We enumerate the corresponding events as $l = 1, 2, ..., n_\lambda$ and $m = 1, 2, ..., n_\mu$, respectively. Within the framework of ES, two events e_l^λ and e_m^μ

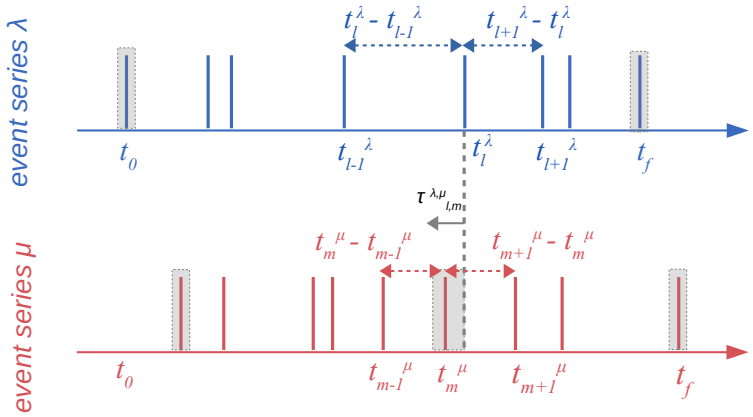

FIGURE 2.2.: Schematic illustration of event synchronization.

are considered synchronized if and only if they both took place with a mutual time difference smaller than a specific, data-adaptive interval $\tau_{l,m}$. In this thesis, we refer to this quantity as a dynamical coincidence interval. The dynamical nature of this time interval implies that the more frequent the events occur (i.e., the shorter the inter-event or waiting time between subsequent events of the same type), the smaller this dynamical coincidence interval is.

$$\tau_{l,m}^{\lambda,\mu} = \frac{1}{2}\min\{t_{l+1}^\lambda - t_l^\lambda, t_l^\lambda - t_{l-1}^\lambda, t_{m+1}^\mu - t_m^\mu, t_m^\mu - t_{m-1}^\mu\}. \tag{2.6}$$

With this definition, we can assess for each pair of events e_l^λ and e_m^μ in the two event series whether or not they have occurred in close succession in comparison with their respective dynamical coincidence interval $\tau_{l,m}^{\lambda,\mu}$. Thereby, we can simply count the number of times an event in series λ is close to a previously occurred event in series μ (and vice versa) as

$$q_{(\lambda|\mu)} = \sum_{l=1}^{n_\lambda}\sum_{m=1}^{n_\mu} c_{l,m}^{\lambda,\mu} \tag{2.7}$$

with

$$c_{l,m}^{\lambda,\mu} = \begin{cases} 1, & \text{if } \tau_{min} < t_l^\lambda - t_m^\mu < \tau_{l,m}^{\lambda,\mu}, \\ 0.5, & \text{if } \tau_{min} = t_l^\lambda - t_m^\mu, \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

Then we can measure the strength of event synchronization between two nodes in a symmetric fashion, where the temporal order of events does not matter,

$$Q_{(\lambda|\mu)}^{ES} = \frac{q_{(\lambda|\mu)} + q_{(\mu|\lambda)}}{\sqrt{n_\lambda n_\mu}}, \tag{2.9}$$

$Q_{(\lambda|\mu)}^{ES} = 1$ implies complete synchronization between two sequences. Fig. (2.2) systematically illustrates the general concept of ES.

Note that, if we are interested in identifying directed influences among the units of a networked system, we might however solely consider the directed quantity $Q_{(\lambda|\mu)}^{ES}$, thereby acknowledging that $Q_{(\lambda|\mu)}^{ES} \neq Q_{(\mu|\lambda)}^{ES}$ is possible. In this thesis, we employ symmetric version of $Q_{(\lambda|\mu)}^{ES}$ to infer possible interconnection between components of a

complex network.

2.1.2.2. Event Coincidence Analysis (ECA)

As described above, the ES method utilizes a dynamical and data-adaptive coincidence interval. Even though this has the advantage of not choosing a specific value a priori, it implies that ES is beneficial if there is some typical inter-event waiting time in the two series to be compared. This might be presumed in the case of EEG recordings exhibiting relatively regular inter-spike intervals, but not necessarily in the context of successions of, e.g., natural disasters which may occur clustered in time. Further, the ES method does not require explicitly determining coincidence interval, which may cause merging information from different timescales and make it difficult to distinguish certain processes. Moreover, depending on the specific application, there might be prior knowledge of typical time scales of the process under study (e.g., a finite propagation speed of events in a spatially extended system) that might be explicitly considered in the corresponding analysis.

Another recently proposed method, event coincidence analysis (ECA), takes the more general viewpoint of possibly coupled point processes. Similar to ES, it is based on counting the cases in which two events of different types have occurred in close succession [44]. However, unlike ES, closeness is defined here in a prescribed static coincidence interval ΔT. We note that, most previous applications of ECA have explicitly acknowledged the fact that the number of coinciding events depends on which series is taken as the reference (which is shared by the ES), thereby distinguishing between the trigger and precursor coincidence rates [44]. In addition to the static coincidence interval ΔT, in its general formulation ECA also considers a second external parameter, the time delay τ, which allows for studying systematic mutual delays between events of two types. Notably, this effect can also be straightforwardly implemented into ES by shifting all event times in one of the series to be compared by a fixed value of τ.

With the aforementioned presumptions, for event sequence λ taken as a reference, the fraction of events in this series that have been preceded by at least one event in the second series μ within a given time window ΔT is referred to as the precursor event coincidence rate between the two event series (see Fig. (2.3)):

$$r_{\lambda|\mu}(\Delta T, \tau) = \frac{1}{n_\lambda - n_\lambda^{'}} \sum_{l=1}^{n_\lambda} \Theta \left[\sum_{m=1}^{n_\mu} 1_{[0,\Delta T]}((t_l^\lambda - \tau) - t_m^\mu) \right]. \tag{2.10}$$

Here Θ represents the left-continuous Heaviside step function preventing events from being counted twice and 1_I indicates the indicator function of the interval I :

$$1_I(x) = \begin{cases} 1, & \text{if } x \in I \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

For a correct normalization of the precursor event coincidence rate, we subtract the total number of events occurring at $[t_0, \tau + \Delta T + t_0]$ interval $(n_\lambda^{'})$ from the total number of events (n_λ). Simply put, due to the nonzero time lag τ, it is not possible for events at λ occurring at $[t_0, \tau + t_0)$ to coincide with any event at μ. Sequentially, considering the events in μ as the basis for normalization, we can define the trigger event coincidence rate as follows:

$$r_{\mu|\lambda}(\Delta T, \tau) = \frac{1}{n_\mu - n_\mu^{'}} \sum_{m=1}^{n_\mu} \Theta \left[\sum_{l=1}^{n_\lambda} 1_{[0,\Delta T]}((t_m^\mu - \tau) - t_l^\lambda) \right]. \tag{2.12}$$

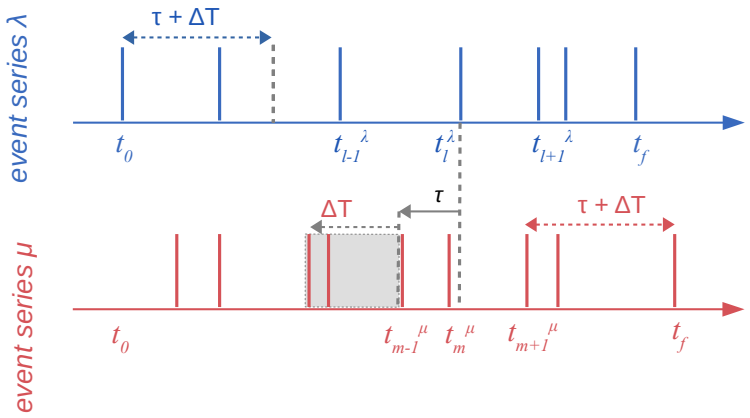

FIGURE 2.3.: Schematic illustration of event coincidence analysis

Correspondingly, for the trigger event coincidence rate we exclude those events that occur at $[t_f - (\tau + \Delta T), t_f]$ interval to avoid possible errors. Now we can define a symmetric matrix of pairwise coincidence rate based on the maximum or the mean value of the two directed trigger events:

$$Q_{\lambda|\mu}^{ECA,Max} = max(r(\lambda|\mu;\Delta T), r(\mu|\lambda;\Delta T)) \tag{2.13}$$

$$Q_{\lambda|\mu}^{ECA,Mean} = \frac{(r(\lambda|\mu;\Delta T) + r(\mu|\lambda;\Delta T))}{2} \tag{2.14}$$

Here $Q_{\lambda|\mu}^{ECA,Max}$ is preferred for highlighting bidirectional dependencies. In contrast, $Q_{\lambda|\mu}^{ECA,Mean}$ is employed when one needs to determine a strong unidirectional connection between two event time series.

2.1.2.3. Comparison Between Pearson's Correlation Coefficient and Event Synchronization and Event Coincidence Analysis

To demonstrate the difference between the performance of the PCC analysis, ES and ECA approaches, we produce three different binary time series with length t = 10000 and 100 independently and uniformly randomly chosen components equal to 1. In this setup, time series x_1 and x_2 are considered to be independent, while event series x_3 and x_1 are dependent. To construct the dependency between x_3 and x_1, we shift non-zero components of x_1 event series forward in time by a time point randomly selected from the set $\{1, ..., 8\}$. Therefore, events in x_3 follow events in x_1 with some time window. Then we attempt to quantify the similarity between x_1 and x_2 using the three methods mentioned earlier. This procedure is repeated 1000 times and the resulting histograms of the corresponding PCC, ES, and ECA values can be seen in Fig. (2.4). According to Fig. (2.4)a, all the three approaches show no strong relationship between events in x_1 and x_2, as we expected. However, Fig. (2.4)b reveals that PCC analysis can not capture the lagged relationship between x_1 and x_3, where there is an unfixed delay between events in two different time series, while values of ES and ECA are considerably higher. We refer to chapter 3, where we pursue a comprehensive numerical study for inter-comparing the potentials of ES and ECA to identify the underlying network structure based on event-type dynamics.

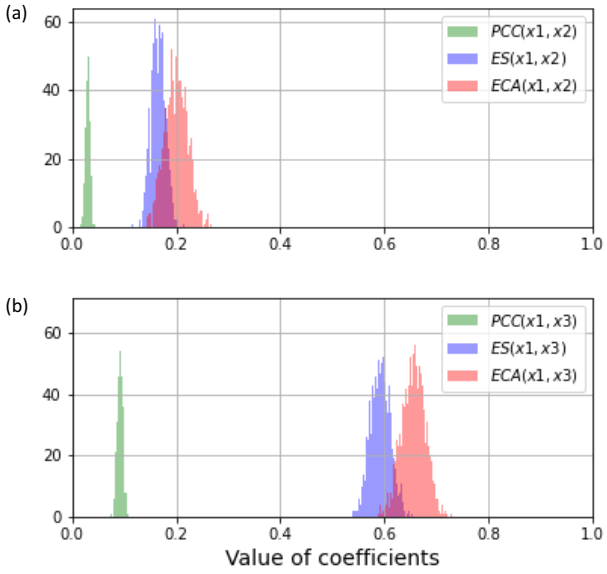

FIGURE 2.4.: Comparison between the performance of Pearson's correlation coefficient (PCC), Event Synchronization (ES) and Event Coincidence Analysis (ECA) for binary data. Here we produce three different event series where x_1 and x_2 are independent, while events in x_1 are followed by events in x_3 within 8 time steps.

2.2. Fundamentals of Stochastic Modelling of Dynamical Systems

The dynamics of complex systems can be represented by different forms of mathematical models [56, 69]. In the last decades, a wide range of mathematical modelings that differ in several aspects (i.e., level of detail or the approximations) has been introduced. Among them, differential equations (e.g., ordinary and stochastic differential equations) are by far the most preferred mathematical tool for modeling real-world phenomena. For instance, deterministic ordinary differential equations (ODEs) have broad applications in fluid mechanics, quantum physics, and statistical physics [109, 128]. An ODE allows us to estimate the evolution of a state variable using the information of its previous states. Typically, for any set of parameters and initial conditions, the solution for a deterministic differential equation is unique and can predict the future perfectly because ODE contains no stochastic elements [86]. However, complex systems are often governed by nonlinear interactions and intricate fluctuations. Therefore, the assumption that a deterministic differential equation can fully explain real-world data is often unrealistic.

A typical approach is then to focus on the comparably few observed, macroscopic variables, assuming that they determine the key dynamics of the system, while the remaining ones are represented by noise. This leads to an approximate, inverse modelling of such systems in terms of Stochastic Differential Equations (SDEs). In the following, we introduce and review three different examples of SDEs (i.e., Langevin equation, Generalized Langevin equation, and Empirical Model Reduction) that are widely used to identify and reconstruct the time evolution of complex systems [59, 108]. Further, we discuss how to extract different terms of these equations from stochastic time series data by employing other data analysis techniques.

2.2.1. Langevin Equation (LE)

In the early twentieth century, Paul Langevin [116] proposed a quantitative description of the random motion of colloidal particles suspended in a fluid, known as a critical problem of non-equilibrium statistical mechanics. This theory's applicability later has been extended to express the dynamical behavior of varieties of macroscopic systems without genuine particle ontologies. The original Langevin Equation is a first-order differential equation, which represents the time evolution of a subset of degrees of freedom containing both frictional and random forces that are associated with the fluctuation-dissipation theorem (FDT) [111]. This theorem implies a systematic internal relationship between the magnitude of the friction and the strength of fluctuating forces.

Consider a system for which the evolution of the macroscopic states $x(t)$ obeys the following equation of motion:

$$\frac{dx(t)}{dt} = a(x,t) + b(x,t)\eta(t). \tag{2.15}$$

where $a(x,t)$ and $b(x,t)\eta(t)$ represent the deterministic force (e.g., friction and gravity) and stochastic forces (e.g., noise and chaotic particle interactions in many-body systems), respectively. Here, $\eta(t)$ is conventionally a stationary, δ-correlated Gaussian process with zero mean: $< \eta(t) >= 0$ and $< \eta(t)\eta(t') >= \delta(t-t')$. The presence of δ-correlated noise indicates that the Langevin process is a Markov process.

Stochastic processes can be also viewed from a different perspective, in terms of the evolution of their conditional probability density function $p(x,t|x',t')$. If a single particle's motion is governed by the Langevin equation, its probability density can be developed according to the Fokker–Planck equation (FPE) [97].

$$\begin{aligned}\frac{\partial}{\partial t}p(x,t|x',t') = \frac{\partial}{\partial x}f(x,t)p(x,t|x',t') \\ + \frac{\partial^2}{\partial x^2}g(x,t)p(x,t|x',t'),\end{aligned} \tag{2.16}$$

Where $p(x,t|x',t')$ refers to the probability of a system to be found in state x at time t. In Eq. 2.16, terms subject to the first and second derivatives are known as drift and diffusion terms, which can depend on position and time. In practice, the drift term determines the deterministic part (slow macroscopic variables) of the underlying dynamics, while the diffusion term reflects fluctuations. One common way to derive the FPE associated with a Langevin process is the Kramers-Moyal (KM) expansion [58, 59]. The KM terms are constructed from a Taylor series of the fundamental master equation of a system which describes the probability evolution of a Markov process [203]. The KM expansion is also called the Generalized FPE since by truncation of the KM expansion after two terms, one can recover the FPE.

For a Markov process, the evolution of the conditional probability densities given by the KM equation is as follows:

$$\frac{\partial}{\partial t}p(x,t|x',t') = \sum_{j=1}^{\infty}\left(-\frac{\partial}{\partial x}\right)^j D_j(x)p(x,t|x',t'), \tag{2.17}$$

where the $D_m(x)$ denotes the mth-order KM coefficients that can be defined by conditional moments $M^j(x,\tau)$ of the variable x and at time τ:

$$\begin{aligned} D_j(x) &= \frac{1}{j!}\lim_{\tau\to 0}\frac{M^j(x,\tau)}{\tau} \\ &= \frac{1}{j!}\lim_{\tau\to 0}\frac{1}{\tau}\langle (x(t+\tau)-x(t))^j \mid_{x(t)=x}\rangle . \end{aligned} \tag{2.18}$$

The beauty of this approach is that one can estimate KM coefficients $D_j(x)$ directly from experimental data by computing transition probability densities in the limit of $\Delta t \to 0$ (numerically, this represents the shortest time increment in the data):

$$D_j(x) \approx \frac{1}{j!}\frac{1}{\Delta t}\langle (x(t+\Delta t)-x(t))^j \mid_{x(t)=x}\rangle . \tag{2.19}$$

Here $\langle (x(t+\Delta t)-x(t))^j \mid_{x(t)=x}\rangle$ represents the mth conditional moment M^m. In this study, to calculate conditional moments from time-discrete data, we employed the non-parametric Nadaraya-Watson estimator [155, 213] which relies on convergence in probability:

$$\frac{1}{\Delta t} < (x_{(i+1)\Delta t} - x_{i\Delta t})^j|_{x_i=x} >= \frac{\sum_{i=1}^{n} K(\frac{(x_{i\Delta t}-x)}{h})(x_{(i+1)\Delta t} - x_{i\Delta t})^j}{\sum_{i=1}^{n} K(\frac{(x_{i\Delta t}-x)}{h})\Delta t}. \tag{2.20}$$

Here, to calculate $D_j(x)$, we assign each data point in the state space to a kernel density and then take a weighted average over all data points. Note that, in the Nadaraya– Watson kernel framework, we need to estimate the kernel and the smoothing parameter called bandwidth. In this study, the kernel function K is assumed to be Gaussian.

It has been demonstrated that the performance of the Nadaraya– Watson kernel estimator largely depends on the smoothing parameter that controls the trade-off between goodness-of-fit and model complexity [48]. Hence choosing an incorrect bandwidth value can lead to an undesirable transformation of the density plot. For instance, a large bandwidth increases the bias by over-smoothing the curve, while a small bandwidth results in a rough estimation. Various methods have been introduced to determine optimal bandwidth, such as the rule of thumb, unbiased and biased cross-validation, and direct plug-in (DPI) [31, 96]. In this thesis the bandwidth h is determined using Silverman's rule of thumb: $h = 1.06\sigma N^{\frac{-1}{5}}$ where σ is the standard deviation of the time series under investigation.

As mentioned earlier, for a continuous diffusion process, there is equivalence between the LE and the FPE descriptions. Therefore we can substitute the drift and the diffusion functions in Eq. 2.15 by the first $D_1(x)$ and second $D_2(x)$ KM coefficients. In other words, the drift and the diffusion terms are approximated by average displacement and conditional variance over time interval τ. Using numerical discretization, in Itô 's interpretation of stochastic integration (see Appendix B.1), the time derivative of the system's trajectory can be written as follows:

$$dx(t) = D_1(x,t)dt + \sqrt{D_2(x,t)}dW(t), \tag{2.21}$$

where dW denotes the increments of a Wiener process (known as a stochastic process with stationary independent normally distributed increments). In this thesis, the numerical integration of the Eq. 2.21 is implemented by the Euler-Maruyama scheme [142]

which is similar to the Euler scheme but also can deal with the stochastic term:

$$x(t+\Delta t) = x(t) + D_1(x,t)\Delta t + \sqrt{D_2(x,t)\Delta t}\eta(t). \tag{2.22}$$

Moreover, the higher-order terms in the KM expansion provide a distinguished test for the continuity of the underlying process. According to Pawula's theorem [168], a stochastic process is statistically continuous and is driven by Gaussian noise if the third and all higher-order KM coefficients tend to zero. Accordingly, three possible processes can be classified in terms of KM coefficients. I) Deterministic processes; where the KM expansion stops at $n = 1$. II) The KM equation reduces to the Fokker–Planck equation when the KM coefficients with order $n > 2$ are negligible. III) For non-vanishing KM coefficients with $n > 2$, the system cannot be accounted as a continuous diffusion process, and the expansion may contain an infinite number of terms. In this circumstance, to reconstruct the relevant aspect of a system in the presence of discontinuous jumps, classical KM formalism requires to be revisited.

2.2.1.1. A two-dimensional Diffusion Process

Further in this thesis, we investigate the potential coupling between two well-known proxies of Greenland ice sheets that contain sudden climate transitions. To this end, we utilize the bivariant diffusion process and try to non-parametrically estimate the parameters of the process.

Consider a two-dimensional diffusion process with form:

$$\overbrace{\begin{pmatrix} \mathrm{d}x_1(t) \\ \mathrm{d}x_2(t) \end{pmatrix}}^{\vec{\mathrm{x}}} = \underbrace{\overbrace{\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}^{\vec{a}} dt}_{\text{drift}} + \underbrace{\overbrace{\begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix}}^{\vec{b}} \overbrace{\begin{pmatrix} \mathrm{d}w_1 \\ \mathrm{d}w_2 \end{pmatrix}}^{d\vec{w}}}_{\text{diffusion}} \tag{2.23}$$

Analogous to the one-dimensional setting, the unknown functions in the above equation can be estimated non-parametrically from a bivariate time series $(x_1(t), x_2(t))$ using KM analysis. The higher KM coefficients in a two-dimensional setting are given by [66] :

$$\begin{aligned} D_{m,n}(x_1,x_2) &= \frac{1}{m!n!}\lim_{\tau\to 0}\frac{M_{m,n}(x_1,x_2,\tau)}{\tau} \\ &= \frac{1}{m!n!}\lim_{\tau\to 0}\frac{1}{\tau}\langle \Delta x_1^m \Delta x_2^n|_{x_1(t)=x_1, x_2(t)=x_2}\rangle. \end{aligned} \tag{2.24}$$

Subsequently, the relationship between the functions in Eq. 2.23 and KM coefficients in Eq. 2.24 can be recovered as follows:

$$D_{1,0} = a_1,\ D_{0,1} \ = a_2, \tag{2.25}$$

$$\begin{aligned} D_{1,1} &= b_{1,1}b_{2,1} + b_{1,2}b_{2,2}, \\ D_{2,0} &= \frac{1}{2}\left[b_{1,1}^2 + b_{1,2}^2\right], \\ D_{0,2} &= \frac{1}{2}\left[b_{2,1}^2 + b_{2,2}^2\right]. \end{aligned}$$

2.2.2. Generalized Langevin Equation (GLE)

In the Langevin equation, the dependencies between slow and fast variables are assumed to be negligible due to the separation of time scales. However, this assumption is not always well-grounded; for instance, in Brownian motion, when the mass of particles is comparable with that of the surrounding particles, the random force does not obey white noise behavior anymore. As van Kampen [98] stated, "Non-Markov is the rule, Markov is the exception". In this scenario, employing the Langevin equation does not lead to an appropriate approximation if the time scale of the macroscopic variables is not much longer than that of the microscopic variables. Moreover, it is also necessary to modify the fluctuation-dissipation theorem to consider how the correlation function of fluctuations varies with the memory of the frictional force. Accordingly, Generalized Langevin Equations (GLE) have been proposed to account for long-range correlations and memory effects of complex systems that do not exhibit strong time scale separation. It is shown that the generalized equation fulfills a particular fundamental consistency condition, which links the memory function to the auto-correlation of the stochastic force.

One of the practical tools to derive a GLE is the Mori-Zwanzig (MZ) formalism [38, 77] which has been initially developed in non-equilibrium statistical mechanics for constructing coarse-grained models. The MZ is a projection-based dimension reduction method that redefines a set of ordinary differential equations into a reduced system with a time-independent Hamiltonian as long as the system is close to equilibrium. The MZ formalism assumes that a macroscopic system can be well described by projecting the full microscopic dynamics of a system onto the space of macroscopic variables. In the following, we briefly show how the GLE can be derived using MZ formalism.

Consider a set of observables $q(x,t) = f(\Phi(x,t))$, where $f : R^N \rightarrow R^d$. The time evolution of such observables can be defined in the Liouville form as follows :

$$\frac{dq(x,t)}{dt} = \mathcal{L}q(x,t), \tag{2.26}$$

where $\mathcal{L}$ is the Liouville operator and $q(x,0) = f(\Phi(x,0)) = f(x)$. The solution of Eq. 2.28 can be written:

$$q(x,t) = e^{\mathcal{L}t}q(x,0) = e^{\mathcal{L}t}f(x). \tag{2.27}$$

$e^{\mathcal{L}t}$ is known as the propagator operator and can commute with the Liouville operator. By substituting $q(x,t)$ with $e^{\mathcal{L}t}f(x)$, we can rewrite Eq. 2.28:

$$\frac{de^{\mathcal{L}t}f(x)}{dt} = e^{\mathcal{L}t}\mathcal{L}f(x). \tag{2.28}$$

To construct the reduced-order representation of a system from N components to d components, MZ exploits a projection operator $\mathcal{P}$, which maps the general system variables onto the subspace of the resolved observables. Subsequently, a projection onto an orthogonal subspace is defined as $\mathcal{Q} = I - \mathcal{P}$ which is the complement of the projection operator $\mathcal{P}$ and satisfies $\mathcal{QP} = \mathcal{QP} = 0$. Applying the identity operator $(\mathcal{Q} + \mathcal{P})$, the Eq. 2.28 is converted into :

$$\frac{d(e^{\mathcal{L}t}f(x))}{dt} = e^{\mathcal{L}t}\mathcal{PL}f(x) + e^{\mathcal{L}t}\mathcal{QL}f(x). \tag{2.29}$$

Using the well-known Dyson Identity [87]:

$$e^{\mathcal{L}t} = e^{(\mathcal{P}+\mathcal{Q})\mathcal{L}t} = e^{\mathcal{Q}\mathcal{L}t} + \int_{t_0}^{t} ds\ e^{\mathcal{L}(t-s)}\mathcal{PL}e^{\mathcal{Q}\mathcal{L}s}, \tag{2.30}$$

we can rewrite the Eq. 2.29 in the following form:

$$\frac{d(e^{\mathcal{L}t}f(x))}{dt} = e^{\mathcal{L}}\mathcal{P}\mathcal{L}f(x) + e^{\mathcal{Q}\mathcal{L}t}\mathcal{Q}\mathcal{L}f(x) + \int_{t_0}^{t} ds\ e^{\mathcal{L}(t-s)}\mathcal{P}\mathcal{L}e^{\mathcal{Q}\mathcal{L}s}\mathcal{Q}\mathcal{L}f(x). \tag{2.31}$$

Different terms on the right-hand side of Eq. 2.31 have well-known interpretations. The first term is a Markovian term that refers to the self-interactions of the macroscopic variables. The second term is related to unresolved variables in the orthogonal space $F(x,t)$, which often is regarded as random noise. Finally, the last term describes memory dependencies between the observable and $F(x,t)$. The resulting lower-dimensional model is known as the generalized Langevin equation that is described as follows:

$$\frac{d}{dt}f(x,t) = \Omega f(x,t) - \int_{t_0}^{t} \mathcal{K}(t-t')f(t')dt' + F(x,t). \tag{2.32}$$

Note that the functional form of MZ may vary depending on the choice of projection operators. For instance, the Chorin's Projection [34] provides the evolution equation for a conditional mean of f while applying the Mori's projection [226] yields the evolution of temporal auto-correlation. In this thesis, we focus on MZ formulation based on the Mori's finite rank projection operator, which is given by:

$$\mathcal{P}h(f(x)) = \sum_{i,j=1}^{d} < h, f_i >< f_i, f_j >^{-1} f_j(x). \tag{2.33}$$

So far, we have seen the MZ formulation in the continuous-time form. However, to construct the evolution equation of a system from empirical data where the outputs are discrete-time snapshots, we require the discrete counterpart of MZ formalism [125, 158]:

$$f((n+1)\Delta t) = \Omega_{\Delta t} f(n\Delta t) - \sum_{k=1}^{n} \mathcal{K}^{k}_{\Delta t} f((n-k)\Delta t) + \xi_{n+1}(x), \tag{2.34}$$

where Ω and $\mathcal{K}$ represent the deterministic function and memory kernel, respectively and can be obtained directly from collected data. Using the evolution of two-time correlation function, we can calculate the Markov matrix $\Omega = \dot{C}(0).C(0)^{-1}$ and the memory kernel [124, 224] as follows:

$$\mathcal{K}^{n}_{\Delta t} = (C((n+1)\Delta t) - \sum_{l=0}^{n-1} \mathcal{K}^{l}_{\Delta t} C((n-l)\Delta t))C^{-1}(0) \tag{2.35}$$

Considering the orthogonal association between the basis function f(0) and the unresolved variables, the noise term is eliminated from the evolution equation of C. After obtaining the memory kernel and the Markov term, we can readily calculate the orthogonal term.

2.2.3. Empirical Model Reduction (EMR)

Over the past few years, various linear and nonlinear inverse stochastic modeling approaches have been intensively developed and applied to obtain reduced models that can explain the statistics of a full system [3, 169]. One commonly used type of inverse stochastic model is the Linear inverse model (LIM). The LIM assumes that the relevant

dynamics can be decomposed into a linear deterministic term and a stochastic white noise process given by:

$$dx_i = L_{ij}x_j dt + d\eta(t), \tag{2.36}$$

where L is a $D \times D$ dynamical matrix and η is a vector of Gaussian white noise. This approach allows modeling dynamical systems in which the fast variables and nonlinear interactions are so fast that they can be approximated by Gaussian white noise.

Although linear simplification (i.e., linearity and stable dynamic) can be considered a proper tool for describing dynamical systems, it may lead to inaccurate results in the presence of nonlinearity and serial correlations. To overcome this drawback, the empirical model reduction (EMR) [110] has been proposed as a nonlinear generalization of LIMs. It has been shown that the EMR with quadratic nonlinearity and additive noise can adequately capture underlying properties of the whole system's variability and can be used later for predictive purposes [30]. For instance, in the context of climate systems, it has been demonstrated that the EMR can successfully model Madden-Julian Oscillation [106] and El Niño-Southern Oscillation [110]. In this thesis, we employ the EMR to reconstruct the dynamics of various synthetic and real-world systems from available time series.

The EMR approach has the following form:

$$dx_i = (A_{ijk}x_j x_k + L_{ij}x_j + c_i)dt + r_{0,i}dt, \tag{2.37}$$

where matrices A and L describe quadratic self-interactions and linear dissipation processes. Here c_i represents intercept vector. Typically to estimate the coefficients in Eq. 2.37, multiple linear regression is used, which attempts to minimize the difference between the observed variables and the model outcome. To account for correlated noise, the EMR models the residual at each level as a linear function of x_i and r_i, given by :

$$\begin{aligned}
\frac{dr_{0,i}}{dt} &= b_i^1\left[x, r\right] + r_{1,i} \\
\frac{dr_{1,i}}{dt} &= b_i^2\left[x, r, r_1\right] + r_{2,i} \\
&\vdots \\
\frac{dr_{l,i}}{dt} &= b_i^{l+1}\left[x, r, r_1, r_2, ..., r_l\right] + r_{l+1,i}.
\end{aligned} \tag{2.38}$$

The presence of these hidden variables that explicitly depend on the past values of slow variables brings forth "memory" effects. An optimal number of levels l can be determined by considering specific stopping criteria. The basic idea is that once the auto-correlation of the last level of residual noise approaches zero and is well approximated by a spatially correlated white-noise process, the process of adding more levels can be terminated. In other words, the residual at the last level in Eq. 2.38, is assumed to obey the Wiener process.

2.3. Fundamentals of Artificial Neural Networks

The rapid growth of data production/acquisition and advancements of computing power brought us closer to understanding complex systems. For instance, in climate-related problems, having access to vast amounts of data gathered from satellites, numerical climate models, and stationary measurement units allow us to extract important spatiotemporal interaction among climate variables through constructing data-

driven models [21, 189]. In recent years, machine learning strategies have emerged as a powerful tool to distill information and make predictions from piles of raw data, besides other data-driven techniques [179]. Machine learning algorithms aim to instruct computers to model a system directly without relying on the driving equations. The two main types of machine learning techniques are supervised, and unsupervised learning, which are applied in different problems with different datasets [5, 54, 194]. Supervised learning aims to teach models to produce desired outcomes based on a labeled training set. In this scenario, we have prior knowledge of what the output should look like. When conducting supervised learning, the main consideration is to develop a model that can determine a function, which maps the input variable (x) to the output observable (y), by minimizing a loss function. On the other hand, unsupervised learning finds inherent structures and patterns without supervision, e.g., principal component analysis [118]. This study focuses on predicting the evolution of complex systems using a form of supervised learning called artificial neural networks (ANNs).

ANNs are information-processing paradigms, inspired by a human brain's structure, designed to mimic the brain functions, such as recognition, classification, perception, and reasoning [129, 179, 212]. The first version of ANNs, known as the Pitts neuron model (MCP), was created in 1943 to model biological neurons, building brain units, using electrical circuits. Later in 1950, with the advent of computers, it was possible to simulate a hypothetical neural network that allows neurons to learn the function that maps inputs into output. However, computer processors were inferior in numerical calculations, logical inference, and data storage compared to the human brain. Due to the recent advance of "big data" and increasing computational power, ANNs have become the center of a technological revolution in many disciplines [179].

Analogous to interconnection in the human brain's neurons, ANNs utilize artificial cells arranged in a layered structure that communicates through weighted connections. There are several different architectures of ANNs, including Feed-forward Neural Networks (FFNNs) [14], Recurrent Neural Networks (RNNs) [67, 146], and Convolutional Neural Networks (CNNs) [160]. One of the most common and well-known ANNs is FFNNs, where the information flows in only one direction from input to output. Hence there exist no feedback connections or loops in this architecture. Generally, FFNNs comprise of three different layers (see Fig. (2.5)):

1) an input layer that receives input data.

2) hidden layers where the number of layers depends on the complexity of the function.

3) an output layer.

In FFNN architecture, each node calculates the sum of the weight of the input and passes this sum to a continuous and differentiable activation function to obtain output values. The output of the node (i) in the (j) layer can be calculated as follows:

$$r_i^j = \mathcal{F}(\sum_k^N w_{ik}^j xk + b_i^j) \ . \tag{2.39}$$

Here $\mathcal{F}$ is a nonlinear activation function that allows constructing a complex mapping from input to the output. Nonlinear activation functions are preferable since linear functions can not capture and learn the complex features of input data. There are different possible choices for the activation function; the most commonly used activation functions are sigmoid, logistic function, hyperbolic tangent, and ReLu functions [190]. In the training process, after a sufficiently large number of training cycles, the weights are adjusted to minimize the loss function by reducing the difference between the actual

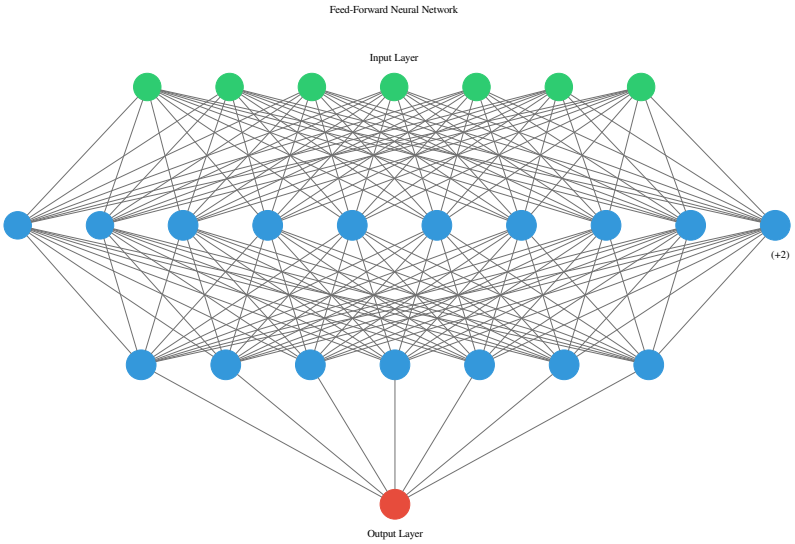

FIGURE 2.5.: A depiction of a simple feed-forward neural network

value and the computed output.

$$\mathcal{L} = \sum_i (y - \hat{y})^2 \, . \tag{2.40}$$

The question is how to minimize the loss function and improve the model performance by changing the parameters of the network's layer (e.g., weights and biases). It can be efficiently evaluated by a supervised learning method called the Backpropagation algorithm (BP) that calculates the gradients of the cost function with respect to the parameters. The key idea is to propagate errors from the output to the input layer and adjust parameters in an iterative process using gradient descent. For instance, consider a FFNN with two hidden layers in Fig. (2.5). First, we try to calculate the gradient of the last weight in the network using the chain rule. Hence we get :

$$\frac{\partial \mathcal{L}_k}{\partial w^{(3)}} = \frac{\partial \mathcal{L}_k}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w^{(3)}} = (y - \hat{y}) r^{(2)} \, , \tag{2.41}$$

Then, we calculate the gradient of layers (2) and (1) as follows:

$$\frac{\partial \mathcal{L}_k}{\partial w^{(2)}} = \frac{\partial \mathcal{L}_k}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial r^{(2)}} \frac{\partial r^{(2)}}{\partial w^{(2)}} \, , \tag{2.42}$$

$$\frac{\partial \mathcal{L}_k}{\partial w^{(1)}} = \frac{\partial \mathcal{L}_k}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial r^{(2)}} \frac{\partial r^{(2)}}{\partial r^{(1)}} \frac{\partial r^{(1)}}{\partial w^{(1)}} \, . \tag{2.43}$$

Eventually, the model's parameters can be updated based on the computed gradient:

$$\Delta w^{(i)} = \gamma \frac{\partial \mathcal{L}_k}{\partial w^{(i)}}. \tag{2.44}$$

Here γ represents the learning rate that scales the magnitude of parameters updates. We note that selecting an optimal learning rate can be challenging. For instance, a too low learning rate requires many updates resulting in slow convergence, while a too high learning rate can interfere with the convergence leading to divergent behavior. Therefore we must find a rate range where the loss function falls sharply.

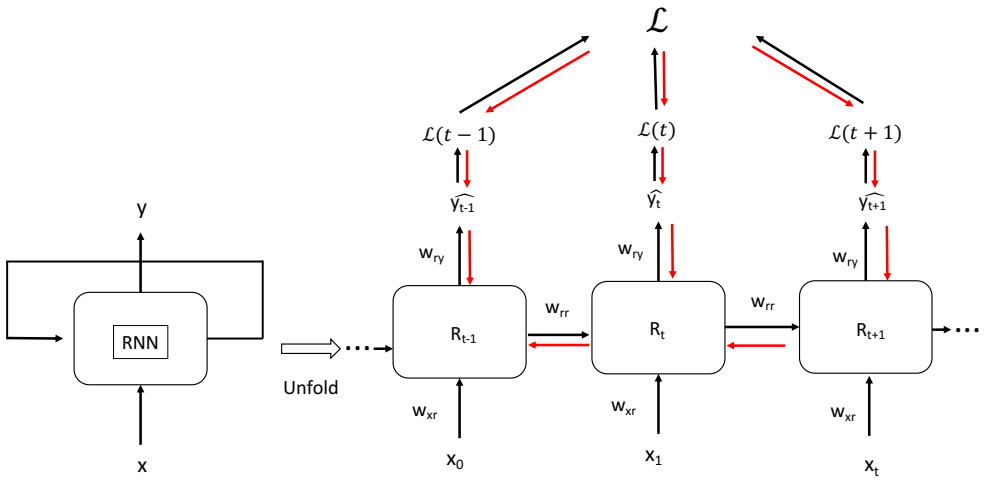

FIGURE 2.6.: A depiction of a single hidden layer Recurrent Neural Network (RNN). The red chain displays the Back-propagation through time (BPTT).

Even though FFNNs are easy to design and are fast, like other techniques, they show some drawbacks. Their performance is sensitive to the initialization of random weights and requires lengthy training data. One of the desirable aspects of modeling a complex system is learning long-term temporal dependencies. However, FFNNs, have no memory of the input. For instance, given a sequence of letters "COMPLEX" when it gets to "L," the FFNN algorithm has already forgotten that it just read "P". Therefore, to learn the temporal behavior, a specific learning framework is required that can preserve past information and keep track of long-term dependencies.

2.3.1. Echo State Networks (ESN)

A specific kind of NNs best suited to interpret time-dependant and sequential data (e.g., climate time series) is Recurrent Neural Network (RNN). Unlike traditional FFNN, where all the inputs and outputs are independent, RNN makes a decision based on the prior information obtained from the sequence of previous inputs. In other words, in the RNN algorithm, the output of the current step becomes the input of the next step allowing the information to persist.

$$r_t = \mathcal{F}(x_t, r_{t-1}; \theta) \ . \tag{2.45}$$

The above equation updates the hidden states using the current input vector $x(t)$ and the context of the previous state $r(t)$. Each hidden state contains information from all states before $r(t)$ for as long as memory can retain. The expression above can be re-arranged in a recursive setting as follows:

$$r_t = \mathcal{F}(x_t, \mathcal{F}(x_{t-1}, \mathcal{F}(..., \mathcal{F}(x_1, r_0; \theta)...; \theta); \theta) \ . \tag{2.46}$$

RNNs exploit the backpropagation through time (BPTT) algorithm to update weights. Conceptually in the BPTT algorithm, errors are first calculated at each time step and accumulated. Then the network is roll-up and parameters are updated [165, 216] (see red chain in Fig. (2.6)).The gradients can be written as a sum of products from the

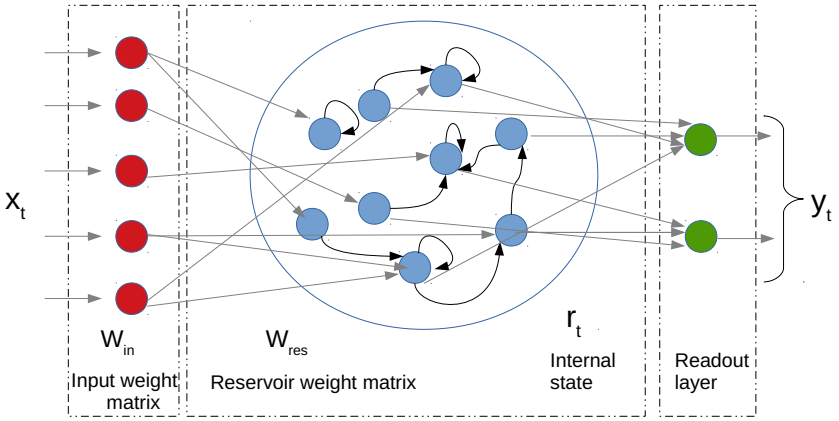

FIGURE 2.7.: A schematic view of Echo State Network

t-th time-step back to the k-th.

$$\frac{\partial \mathcal{L}_t}{\partial \theta} = \sum_{t=1}^{T} \left(\frac{\partial \mathcal{L}_t}{\partial r^{(t)}} \frac{\partial r^{(t)}}{\partial r^{(k)}} \frac{\partial r^{(k)}}{\partial \theta} \right) \\ \frac{\partial r^{(t)}}{\partial r^{(k)}} = \prod_{t>i>k} \frac{\partial r^{(i)}}{\partial r^{(i-1)}}. \tag{2.47}$$

Despite its advantages, we can face two problems during the backpropagation of time-series data, known as the vanishing and exploding gradient problems. The first problem arises when $||\frac{\partial r_{t+1}}{\partial r_t}||_2$ in Eq. 2.47 goes to zero exponentially fast. When the gradient vanishes, updating the model's parameters will no longer be significant in the learning process, making it difficult to capture some long-range dependencies. In the second case, the gradient norm grows exponentially fast during training, ending up with NaNs. To avoid the exploding and vanishing gradients problem, an alternative new paradigm called the Reservoir computing (RC) approach has been introduced [135]. In the traditional RRN algorithm, both the hidden layer and the readout weights get updated during the training phase (see Fig (2.6)). While the reservoir connection weights in RC are randomly generated and kept unchanged. In the RC framework, the training is mainly for the readout part, where only weights of the connections from the reservoir units to the readout are tuned. Since most of the parameters are fixed, RC requires fewer training data, making the learning process quick and stable. The three well-known examples of the RC class are Echo State Networks (ESNs) [89], backpropagation-decorrelation neural networks, and Liquid State machines (LSTM) [197]. In this thesis, we concentrate on the simplest form of RC model, ESN, to predict climate variability.

ESN has been applied in a variety of tasks from classical time-series prediction [120, 167] to language modeling and speech recognition [207] to dynamic pattern classification. ESNs are comprised of three parts: an input layer, a so-called reservoir that processes the input, and an output layer that uses the reservoir output for prediction (see Fig (2.7)). The internal unit can be considered as a high-dimensional dynamical system with states $r(t)$ that evolve according to the following equation:

$$r(t+1) = \mathcal{F}(W_r r(t) + W_{in} x(t)) \ . \tag{2.48}$$

Where $W_{in} \in \mathbb{R}^{N \times L}$ is the input matrix mapping the input of dimension L to the

reservoir space of dimension N. $W_r \in \mathbb{R}^{N \times N}$ is a sparse, weighted adjacency matrix with elements drawn randomly from a normal distribution with zero mean. Here $\mathcal{F}$ indicates a nonlinear activation function for which we choose the hyperbolic tangent. Contrary to most conventional NN models, in the ESN setup, the values of W_{in} and W_r are not optimized but fixed at randomly chosen values. Hence, the only trainable connections are given by the reservoir-to-output layer matrix W_{out}, through which the reservoir states are mapped back to the L-dimensional outputs $y(t)$. Moreover, the reservoir does not require remembering the entire temporal information of the past; hence, it skips unnecessary information after some time, resulting in extremely fast converging of the training. The linear readout can be computed as:

$$y(t+1) = (W_{out} r(t+1)) \; . \tag{2.49}$$

Optimal values for W_{out} that minimize the loss function can be easily determined by simple linear regression. However, traditional linear regression can cause unreliable parameters estimation in ill-posed problems, leading to less generalizability. There are different variants of ordinary least square methods (OLS), e.g., Ridge regression and LASSO [80, 205], that have been suggested to improve the numerical instability and reduce overfilling by introducing a penalty term. This thesis calculates the output unit weights using Ridge regression (also known as Tikhonov Regularization). The Ridge regression cost function can be demonstrated mathematically as:

$$\mathcal{L} = \sum_{t=0}^{T} ||W_{out} r(t) - x_d(t)|| + \lambda ||W_{out}||^2 \; . \tag{2.50}$$

where $||..||$ is the L_2-norm of a vector and λ is a regularization parameter which penalizes larger output weights W_{out}. Accordingly, the regularized optimal output matrix W_{out} is determined via:

$$W_{out} = (R^T R + \lambda \mathbb{I})^{-1} R^T x_d \tag{2.51}$$

Here R refers to the network's states, and $\mathbb{I}$ is the $N \times N$ identity matrix. Then the prediction phase initiates with the information of the reservoir state $r(t^*)$ at the last time step of the training phase to make the prediction $y(t^*+1)$ using the trained output weights W_{out}. Further predictions are then iteratively made forward in time by passing $y(t^*+1)$ to the ESN as input to produce a forecast $y(t^*+2)$, and so on. Therefore, with well-defined inner weights, the reservoir can capture the particulars of the input by developing a high memory capacity.

2.3.1.1. Hyper Parameters of Echo Stat Network

In analogy to other NNs, the predictive performance of the ESN depends on various hyperparameters, such as the choice of the reservoir size (N), the sparsity (p) of W_{in}, and the spectral radius ρ of W_r. Reservoir size N identifies the dimension of the space of the reservoir. Generally speaking, the larger the reservoir, the better the approximation of the underlying dynamics of the system, and the easier it is to find a linear combination of the signals to approximate output. However, an extensive network can be computationally expensive; hence, selecting optimal hyperparameters with a smaller network and later scaling it to the larger one is advised. Sparsity p is another critical parameter in designing a good ESN, which determines the distribution of nonzero elements in the reservoir matrix W_r. This can be implemented by a uniform or normal distribution centered around zero. Sparser networks are preferable to increase the ESN architecture's computational efficiency. We note that the input matrix W_{in}

is usually generated by the same type of distribution as W_r. Another key parameter affecting the ESN performance is the input scaling of the W_{in}, which determines the overall nonlinearity of the reservoir units. In other words, the input scaling dictates how far the network's hidden states are pushed away from the linear part of the activation function [148]. For instance, using hyperbolic tangent ($tanh$) as an activation function, a small input scaling results in the operation of reservoir states in the more linear part of the tanh curve.

One of the most critical parameters is the spectral radius ρ of the reservoir connection matrix W_r. This hyperparameter is the largest absolute eigenvalue of the matrix W_r and scales the width of the distribution of nonzero elements. This parameter controls the echo state properties (ESP) [90, 133, 134] of the reservoir, which indicates the rate at which the history of the input vanishes gradually in time. For a task requiring longer input memory, the closer ρ is to one, the longer the network's capability is to memorize past inputs. It is commonly assumed that ρ must not exceed unity for a reservoir to show ESP. However, Yildiz et al. [222] demonstrated that spectral radius below unity is insufficient to guarantee the ESP. Therefore, there is no general condition for optimal spectral radius, and it must be determined by task-specific experimentation.

2.3.1.2. Comparison Between the ESN and FFNN Performance

Conventional ESN and FFNN have been applied to various nonlinear deterministic systems (e.g. Lorenz system in Fig. (2.8) and data with low level of noise [28]. Here we compare the performance of the ESN and FFNN on the Lorenz system which is a simplified model for atmospheric convection. The dynamics of a Lorenz system can be described by a set of coupled nonlinear equations for the variables x, y and z as a function of time:

$$\begin{aligned} \dot{x}(t) &= \sigma(y - x), \\ \dot{y}(t) &= x(\rho - z) - y, \\ \dot{z}(t) &= xy - \beta z. \end{aligned} \tag{2.52}$$

We generate a long trajectory with 13000 points separated by a time interval Δt = 0.01. Here, we choose the standard set of parameters $\sigma = 10$, $\rho = 28$, and $\beta = \frac{8}{3}$. Fig. (2.8) shows the short-term prediction skills of ESN (solid lines) and FFNN (dashed lines). We see that the ESN predictions follow the true trajectory with high precision for a long time, while the predictions from FFNN diverge from the true state after 200Δt. It can be concluded that the ESN significantly outperforms FFNN in the short-term perdition of sequential data. The x-axis in Fig. (2.8) is in units of the Lyapunov time λ_{max}^{-1}, where λ_{max} is the maximum Lyapunov exponent.

As we mentioned in section 2.3.1.1, the performance of the ESN algorithm substantially depends on the value of hyperparameters. Fig. (2.9) displays the performance of the ESN on forecasting the Lorenz system for hyper-parameters (a) the input scaling, (b) spectral radius, and (c) reservoir size. For the Lorenz system, we could achieve the highest performance for the set (input-scaling= 0.1, $\rho = 0.2$, and $N = 400$). The FNN employed in this study contains three hidden layers, each with 100 neurons. The network weights are optimized accordingly by the Adam optimizer, a stochastic gradient descent algorithm.

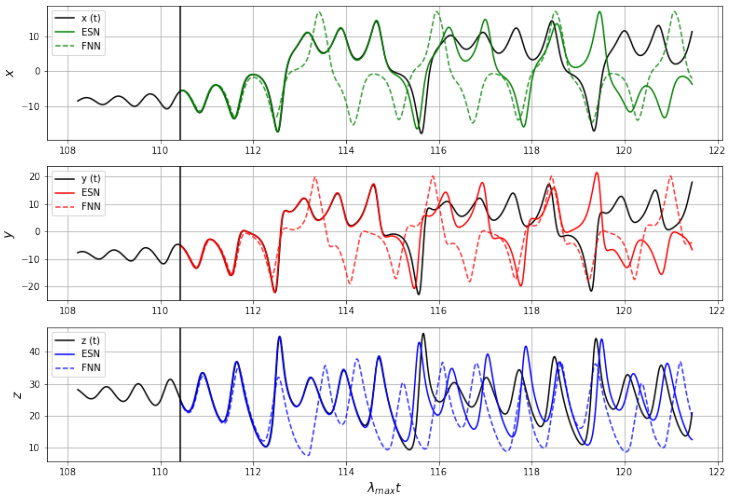

FIGURE 2.8.: Forecasting Lorenz system using the ESN and conventional FFNN algorithms.

2.3.2. Past Noise Forecasting

Most real-world time series are subjected to stochastic forces making it difficult to fully predict the evolution of variables of interest in the presence of noise. In this thesis, to improve the prediction of real-world time series, we propose an efficient technique based on the ESN in combination with a time-series decomposition method to model the high-frequency fluctuations that have posed substantial problems for training neural networks in the past. To model the high-frequency component, we apply the Past-Noise Forecasting (PNF) method [30].

A possible solution is to estimate the time-dependent high-frequency forcing from suitable samples from its history. Chekroun et al. [30] originally developed a particular prediction methodology called Past Noise Forecasting (PNF) to circumvent this problem. The purpose of PNF is to select the best sample of past stochastic forcing to drive the system into the future, using the knowledge of the past noise trajectory. To find such potential noise realizations from past parts of the time series, the stochastic forcing is conditioned on the low-frequency component of the underlying system. They employed the non-parametric Singular Spectrum Analysis (SSA)[74, 209] as a technique to decompose the time series into a low-frequency component and the corresponding high-frequency part. The key idea behind the PNF method is that the model's sensitivity to the initial state exhibits linear or weak response in time scales less than L, and we can assume that the stochastic forcing determines the phase of the system.

Now, consider the time series of high-frequency variability χ_t of a system of interest. One can derive samples of stochastic forcing from the past by splitting the χ_t time series into N continuous fragments of length L each $\{\chi^{t_i} : i = 1, ..., N\}$ with $t_i \in [t^* - L]$ that can potentially drive the system, from arbitrary t^* to $t^* - L$. To select the best noise segment χ^{t_i}, the phase of the system at time t must be inspected. To do so, the low-frequency component (LFC) is split into different segments of length Δ and searched for those segments that resemble the reference LFC segment just preceding

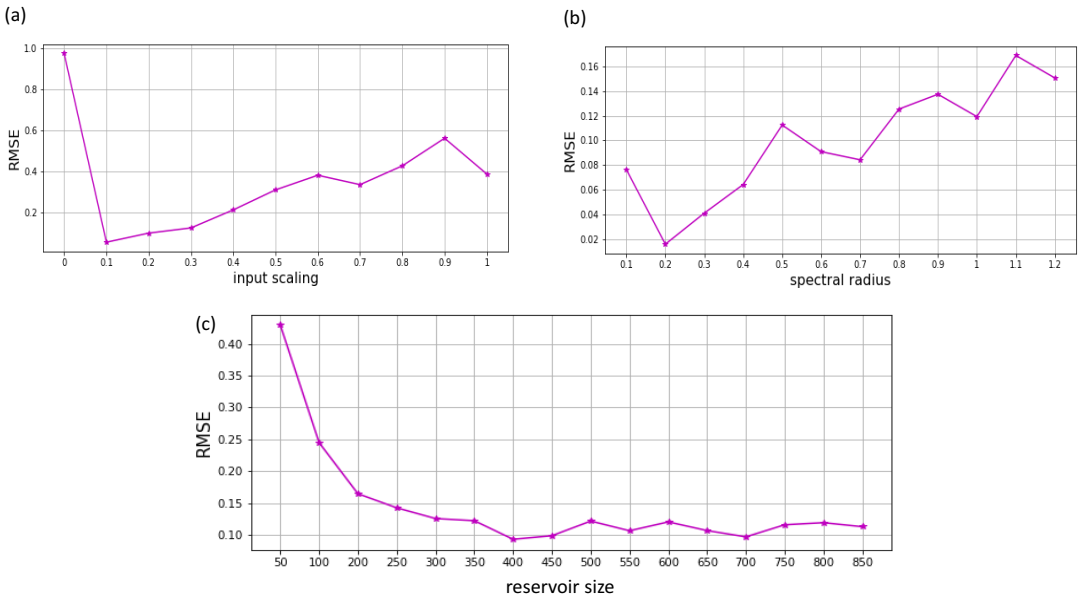

FIGURE 2.9.: Root Mean Square Error (RMSE) of the Lorenz system with respect to variations of **(a)** input scaling, **(b)** spectral radius, and **(c)** reservoir size. The errors are averaged over 100 realizations for each parameter.

time step t^*, at which we initiate the forecast:

$$\begin{aligned} RMSE(LFC(t_j + \Delta, t_j) - LFC(t^* - \Delta, t^*)) \leq \alpha \\ CORR(LFC(t_j + \Delta, t_j) - LFC(t^* - \Delta, t^*)) \geq \gamma \end{aligned} \tag{2.53}$$

In practice, the skill of the PNF method depends on the parameters Δ, choice of K (number of leading reconstructed components that capture the LFC), and L. Hence they must be appropriately tuned. With these selection criteria, one can then identify high-frequency components starting at different $t_j < t^* - \Delta$, which we then use to force the ESN prediction for times $t > t^*$.

$$\Xi_{t^*}(\alpha, \gamma) := \{t_j \in [0, t^* - \Delta]\} \tag{2.54}$$

$$\hat{F}_{t^*} := \{\chi^{t_i} \in F_{t^*} : t_i \in \Xi_{t^*}\} \tag{2.55}$$

Here $\hat{F}_{t^*}$ is the subset of noise segments that can be used to drive the system into the future for times $t > t^*$.

Part II.

Applications

3. Using Event Synchrony Measures for Network Inference From the Timing of Events

One of the most important types of dynamical processes on real-world complex networks are spreading processes in which the effect of a localized incident is propagated throughout the system along the existing connections with a specific probability. This event (or more generally information) propagation process reflects the dynamics of many real-world networks, describing such opinion formation [92], disease outbreak [24], and climate-related extreme events propagation [22]. While the effects of network topology on spreading efficiency have been vastly studied [166], we here address the inverse problem of whether we can infer an unknown network structure from the timing of events observed at different nodes. For this purpose, we numerically investigate two types of event-based stochastic processes. On the one hand, we consider a generic model of event propagation on networks where the nodes exhibit two types of event-like activity: (I) spontaneous events reflecting mutually independent Poisson processes, and (II) triggered events that occur with a certain probability whenever one of the neighboring nodes exhibits any of these two kinds of events. On the other hand, we study a variant of the well-known Susceptible-Infected-Recovered-Susceptible (SIRS) model [83] from epidemiology and record only the timings of state switching events of individual nodes, irrespective of the specific states involved. Based on simulations of both models on different prototypical network architectures, we study the pairwise statistical similarity between the sequences of event timings at all nodes through event synchronization (ES) [174] and event coincidence analysis (ECA) [44] (see section 2.1.2 in chapter 2). By taking strong mutual similarities of event sequences (functional connectivity) as proxies for actual physical links (structural connectivity), we demonstrate that both approaches lead to reasonable prediction accuracy. The results presented and the figures shown in this chapter are based on publication (Hassanibesheli, F., Donner, R. V. (2019). Network inference from the timing of events in coupled dynamical systems. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(8), 083125) with the permission of AIP Publishing.

3.1. Generic Event propagation model

Event sequences are discrete events that occur at specific times and can be regarded as stochastic point processes. The first part of our research is to create a stochastic process model for the propagation of events on a network. Here, we initialize seed events independently and randomly at each node, which results in homogeneous Poisson processes at each node. To simplify the process, we keep the intensity (rate) α of those processes the same for each node. At this stage, the inter-event times at any given node follow an exponential distribution with the same characteristic parameter, while the number of events observed in a given period of time exhibits a Poisson distribution with the same parameter. To model event propagation along the edges of G, we now let each individual event at a given node propagate to any of its neighbors in the network with a certain probability p and time delay τ_p. Fig. (3.1) provides a

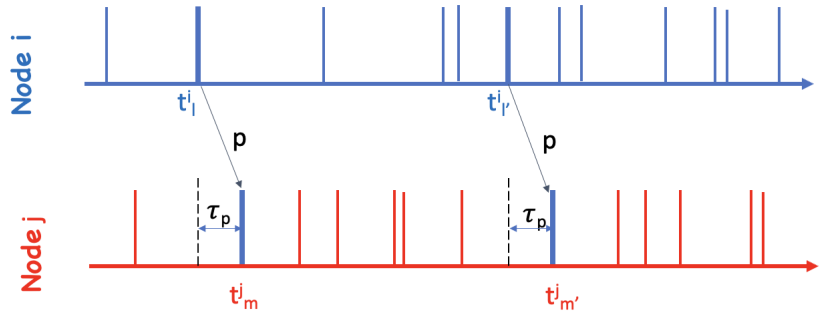

FIGURE 3.1.: Schematic illustration of the propagation process between two series of random events observed at two neighboring nodes i and j. Events can be propagated through the existing link with a prescribed probability p and time delay τ_p. Accordingly, some events at node i are followed by events at neighbor j, while others are not.

schematic illustration of this propagation process. Suppose that at time t, an event takes place at node i. Then, there is a probability p of this event being propagated to any of its neighbors independent of each other, which means that at time $t + \tau_p$, each of the neighboring nodes can experience such a triggered event with probability p. In order to avoid back-and-forth propagation of events between two neighbors, we restrict this propagation to one direction at a time. Hence, the event originally taking place at time t at node i and being propagated from node i to node j, cannot reappear at node i at time $t + 2\tau_p$.

Based upon this construction principle, we can express the appearance of an event at node i and time t by the binary indicator variable S_i^t where it equals 1 if and only if the event takes place. Therefore the probability of an arbitrary node i to experience an event at time $t + \tau_p$ that has been triggered by an event at one of its neighbors at time t can be approximated by:

$$P(S_j^{t+\tau_p} = 1) = 1 - (1 - p)^{\sum_{i=1}^{N} A_{ij} S_i^t}, \tag{3.1}$$

where $A_{ij} = 1$ ($A_{ij} = 0$) describes the existence (non-existence) of a link from i to j. Note that we neglected the probability of emergence of a random seed event at node i (if time is assumed continuous, this would be zero; for discrete-time, we would just have to add a constant) as well as the above-described prohibition of an event traveling back-and-forth along with the same bidirectional link. Moreover, the values of λ and p should be chosen carefully to avoid too frequent events, in which case purely random co-occurrences may be observed too often to still be able to infer the underlying network connectivity from the timing of events anymore. Since the inclination of this study is to work with extreme (rare) event series, we restrict ourselves here to the case of a fixed $\tau_p = 1$ and global $p = 0.4$. It is easy to see that increasing the propagation probability would increase the event rate.

3.2. Network Inference

One important fact to be realized is that the numerical values of both $Q^{ECA}_{\mu|\lambda}(\Delta T, \tau)$ and $Q^{ES}_{\mu|\lambda}$ (see Eq. 2.14 and Eq. 2.9) may have only limited utility for the actual statistical inference of network links, depending on the respective number of events recorded and their associated inter-event time distribution [44]. For the case of ECA, corre-

sponding pairwise statistical significance tests have been discussed in great detail by Donges *et al.* [44]. On the one hand, there exists a simple analytical significance test against the null hypothesis of two sequences of mutually independent Poisson processes, which under the assumption of sufficiently rare events can be formulated as an exact binomial test. In turn, due to the consideration of data-adaptive dynamical coincidence intervals (for definition see Eq. 2.6), such an analytical test does not exist for ES. Therefore, and because the underlying assumptions for the analytical test for ECA are very restrictive, it is recommended to consistently employ a simple form of surrogate-based significance tests making use of shuffled events for estimating the expected distribution of $Q^{ECA}_{\mu|\lambda}(\Delta T, \tau)$ and $Q^{ES}_{\mu|\lambda}$ in the case of independent event sequences. Therefore, we follow the aforementioned concerns regarding the applicability of statistical tests with analytical form to both ES and ECA and focus on the numerical approximation of the respective test statistics (ES strength and event coincidence rate, respectively) based on surrogate data. For this purpose, we randomly shuffle each binary event time series 300 times without replacement (hence, conserving only the number of events in each series, which is justified as long as serial dependencies within each sequence are negligible) and calculate the two similarity measures for each pair of surrogate event series. Then, the distribution of all values from the surrogate ensemble is used as a numerical estimate of the unknown analytical distribution of the test statistic, and the empirical values of ES strength and event coincidence rate for the original sequences are compared with those distributions. Ultimately, we employ a typical confidence level of 0.05 by considering a link between two nodes to be likely present if the similarity coefficient of the original event series is larger than the respective 95th percentile of the test distribution. In other words, when the p-value of the pairwise ES or ECA based similarity measure is smaller than the associated confidence level of 0.05, the synchronization between event series likely cannot be explained by chance.

In summary, we thus obtain an estimate of the unknown adjacency matrix of the underlying network as follows:

$$A_{\lambda|\mu} = \begin{cases} 1, & \text{if } Q^{ECA}_{\lambda|\mu}(\Delta T, \tau) > Q^{*ECA}_{\lambda|\mu}(\Delta T, \tau), \\ 0, & \text{else} \end{cases} \tag{3.2}$$

(respectively, the same for $Q^{ES}_{\lambda|\mu}$), where the superscript * indicates the value of the similarity measure corresponding to the respective confidence level of 0.05 obtained from the surrogates.

3.3. Results and Discussion

We first perform a numerical simulation of the generic event propagation model (see section 3.1) with an initial event rate of seed events $\alpha = 0.006$ and $T = 5000$ time steps (i.e., having a spontaneous event each 165 time steps on average). As a typical network size allowing for efficient numerical analysis, we focus on the case of $N = 50$ unless stated otherwise. Based on the 50 event series resulting from different model runs, we have checked those series' empirical waiting time distributions for their consistency with an exponential distribution. By employing the Kolmogorov-Smirnov test [145] as a goodness-of-fit measure that compares the empirical distribution of the data with a known reference, we find that the observed waiting times of the combined sequences of spontaneous (seed) and triggered events are statistically indistinguishable from an exponential distribution.

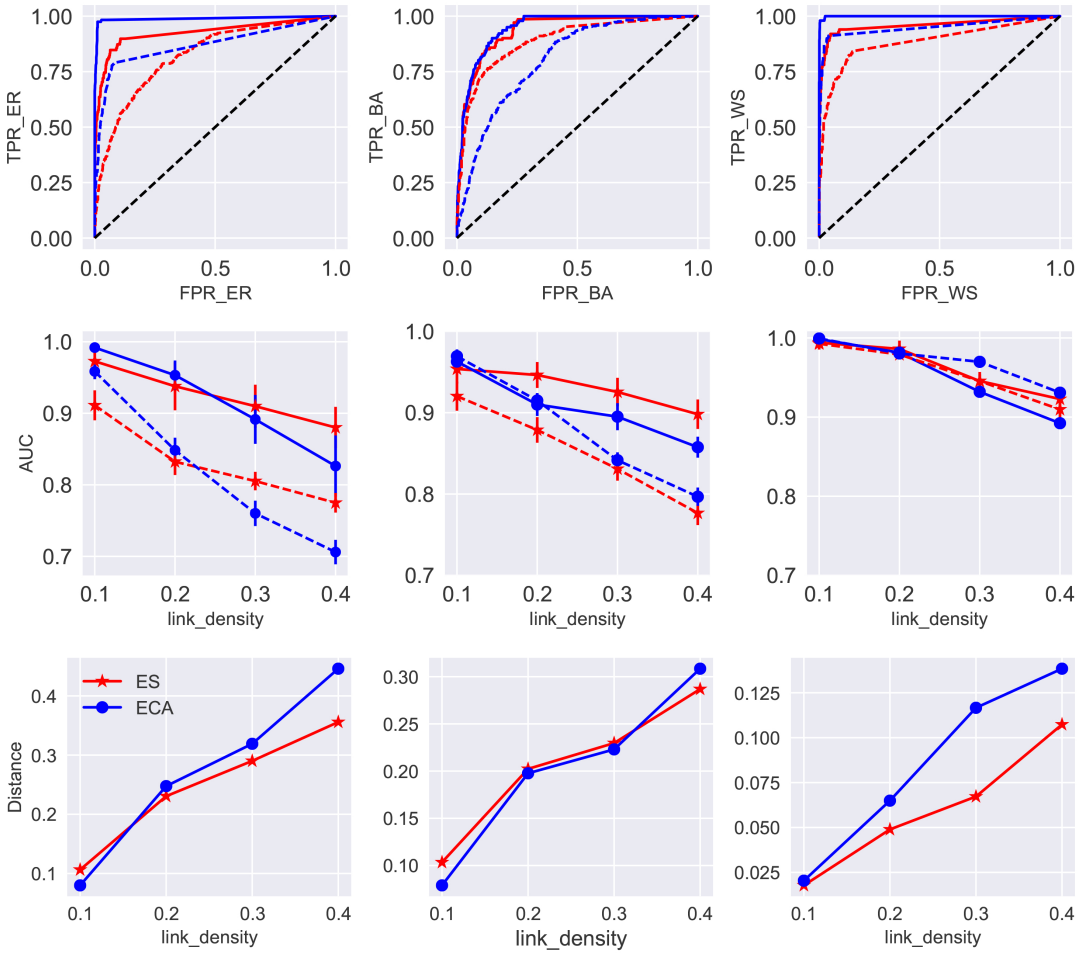

FIGURE 3.2.: ROC analysis characterizing the performance of network inference based on ECA (blue) and ES (red). From left to right, the results correspond to Erdös-Rényi (ER) random graphs, Barabási-Albert (BA) scale-free networks, and Watts-Strogatz (WS) small-world networks (with a rewiring probability of the original underlying ring lattice of $\pi = 0.3$). Top row: ROC curves for randomly chosen realizations of ER networks with link densities of 0.1 (solid lines) and 0.3 (dashed lines), BA networks and WS networks with a mean degree of 5 (solid lines) and 10 (dashed lines), respectively, each with $N = 50$ nodes. Middle row: Dependency of the reconstruction performance as expressed by the area under the ROC curve (AUC) on the link density ρ for the corresponding network topologies for networks consisting of 20 (solid lines) and 50 nodes (dashed lines), respectively. Shown are the mean values and standard deviations obtained from ensembles of 300 simulations for each setting. Bottom row: Minimum distance of the ROC curve from the upper left corner of the ROC diagram as a function of the link density for networks with $N = 50$ nodes.

We now discuss the performance of both ES and ECA in reconstructing the network connectivity based on the observed event dynamics. We emphasize that starting with the directed similarity measures r and q, we have tested two versions of symmetrization based on taking the mean or maximum of both directions between each pair of nodes (see Eq. 2.9 and Eq. 2.14), finding no significant differences between the corresponding results obtained in all situations studied in the following. Accordingly, we will only report the results where the maxima of the respective similarity coefficients in both possible directions are considered. To provide a quantitative comparison between ES

and ECA regarding their performance in the network inference task (discussed in section 3.2), a corresponding evaluation criterion is required. As a corresponding framework that is widely applied in binary classification problems, we utilize here the receiver operating characteristic (ROC) [53, 73] (see Appendix A.2 for more details). This approach allows quantifying how successfully both similarity measures correctly identify the presence of links in the underlying network structure. Hence to determine the optimal threshold for transforming the pairwise similarity matrices resulting from ES and ECA into the binary adjacency matrix, we first consider all possible values and subsequently examine the resulting ROC curves. For each point of the obtained curves, we compute the distance from the upper left corner of the ROC diagram (TPR = 1, FPR = 0) and identify the point with the shortest distance. Here TPR and FPR stand for true positive and false positive rates, respectively. The associated threshold value is then considered the best choice and will be referred to in all following results. To assess the performance of ES and ECA on different structures, we considered three types of random graphs: Erdös-Rényi random graphs, scale-free Barabási-Albert networks [12], and small-world Watts-Strogatz networks [214].

Fig. (3.2) displays the resulting reconstruction performance for different network topologies. Different colors indicate the two similarity measures for the network reconstruction (blue: ECA, red: ES). As demonstrated by the resulting area under the curve (AUC) values, the accuracy of both methods depends on the underlying topology of the network. In all cases, increasing the mean degree (i.e., increasing the link density $\rho = \langle k \rangle / (N-1)$) for networks with 50 and 20 nodes, respectively, generally reduces the AUC. In other words, we obtain a better accuracy of link prediction for sparser networks. This effect could partially result from the rising number of closed loops in the network, which generate events circulating among groups of nodes. To account for the two possible parameters of the ECA, the resulting reconstruction accuracy has been carefully checked for different combinations, revealing the best performance for $\Delta T = 1$ and $\tau = 1$, i.e., a coincidence delay of one time step corresponding to the imposed propagation delay as one should have expected. Both similarity measures successfully identify a vast part of the existing links in all cases. The ECA approach exhibits a tendency towards achieving slightly higher accuracy than ES at sparse networks (for example, in the case of the Erdös-Rényi network with $N = 50$ nodes and a link density of ρ =0.2, the AUC based on ECA is 0.869 while dropping to 0.82 for ES). For denser networks, using ES for link identification commonly results in better network reconstruction, which could be related to the fact that the dynamical coincidence interval of ES captures associations at different time scales rather than a single scale as that fixed by the ECA parameters, thereby considering a larger number of events as coinciding events.

In order to investigate the overall reliability of the respective reconstruction approaches for a single network realization, we study the empirical distributions of TPR and FPR as well as AUC obtained from in total 300 different realization of each model, which are shown in Fig. (3.3). For the specific settings investigated here, ECA shows its benefit for the Watts-Strogatz network as seen by considerably higher AUC and TPR values than for the ES based reconstruction at similar FPR values. In turn, for the Erdös-Rényi networks, both TPR and FPR are larger for the ECA based reconstruction as compared to the results of ES, leading to overall similar AUC values.

Finally, we examine the sensitivity of the obtained results regarding changes in the network density ρ and propagation probability p for the example case of ECA-based network reconstruction applied to an Erdös-Rényi random network with $N = 20$ nodes. Fig. (3.4) shows the corresponding results. First of all, regarding the effect of p, we see that in the absence of triggered events ($p = 0$), the node dynamics exhibits

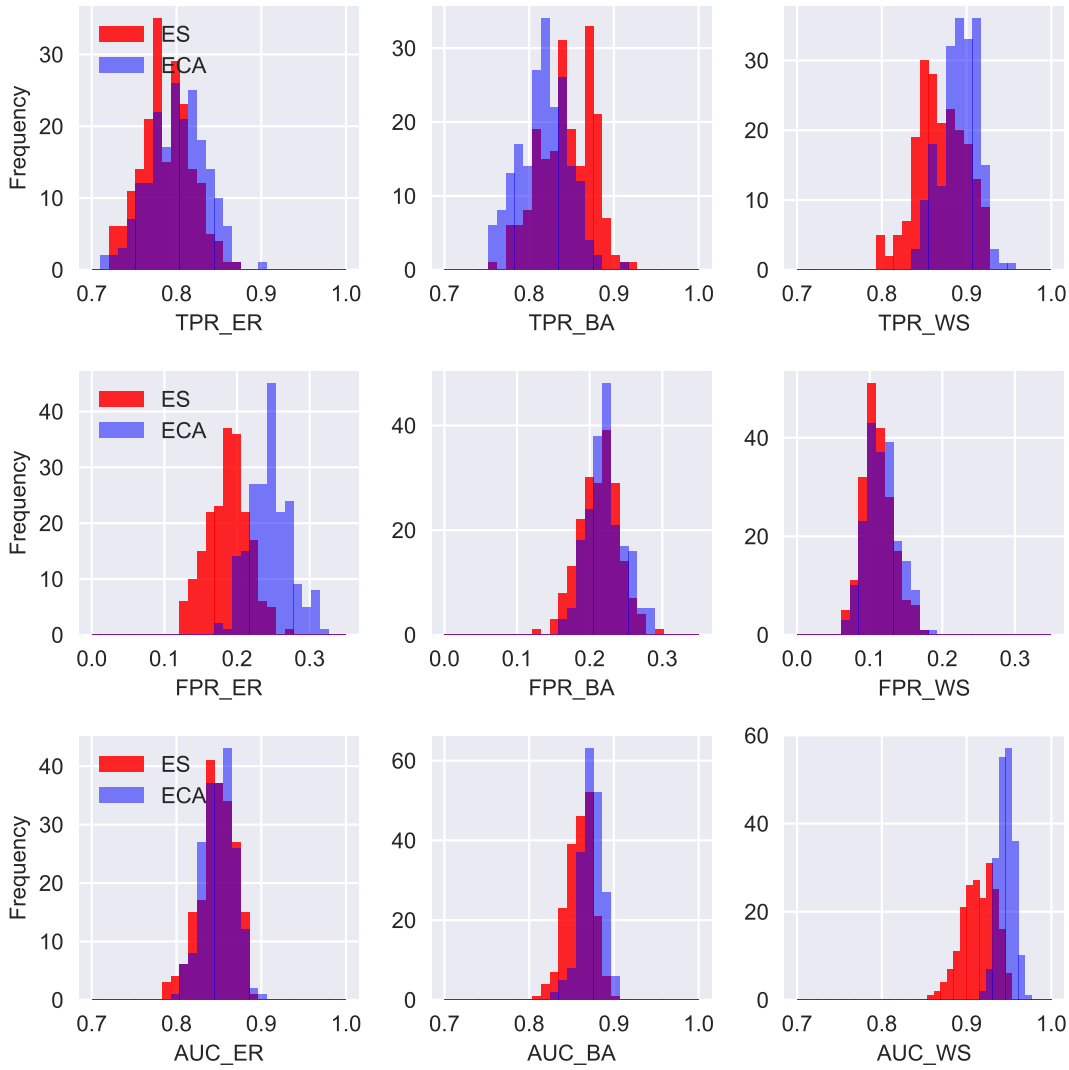

FIGURE 3.3.: Histograms of TPR, FPR and AUC (from top to bottom) for network reconstructions based on ECA (blue) and ES (red) obtained from 300 realizations of the generic event propagation model on Erdös-Rényi, Barabási-Albert and Watts-Strogatz networks with $\pi = 0.3$ (from left to right) with 50 nodes and link densities of 0.2.

pairwise independent Poisson processes, so that the network inference behaves like a random guess (AUC ≈ 0.5) as expected. For very small propagation probabilities, the random seed events still blur the information provided by the triggered events at small network density, thereby leading to an imperfect network inference. However, as soon as sufficiently many events are propagated through the network, the link prediction becomes rather reliable in sparse networks (small ρ) while successively decreasing in performance as the network gets denser. In general, we find a maximum link prediction accuracy at still relatively small yet clearly non-zero propagation probabilities, which is successively shifted towards smaller p for larger link densities. It is interesting to note that, if considering the combined effect of $\langle k \rangle p$, the resulting AUC values approximately collapse to a unique, monotonically decreasing curve (for sufficiently large $p \gtrsim 0.3$). The shape of this curve can be related to the changing character of the considered process at $\langle k \rangle p > 1$, becoming essentially a growing branching process. We only find some deviations from this limiting curve for smaller p and larger ρ, possibly reflecting the finite network size. In general, alternating the network size (e.g., using

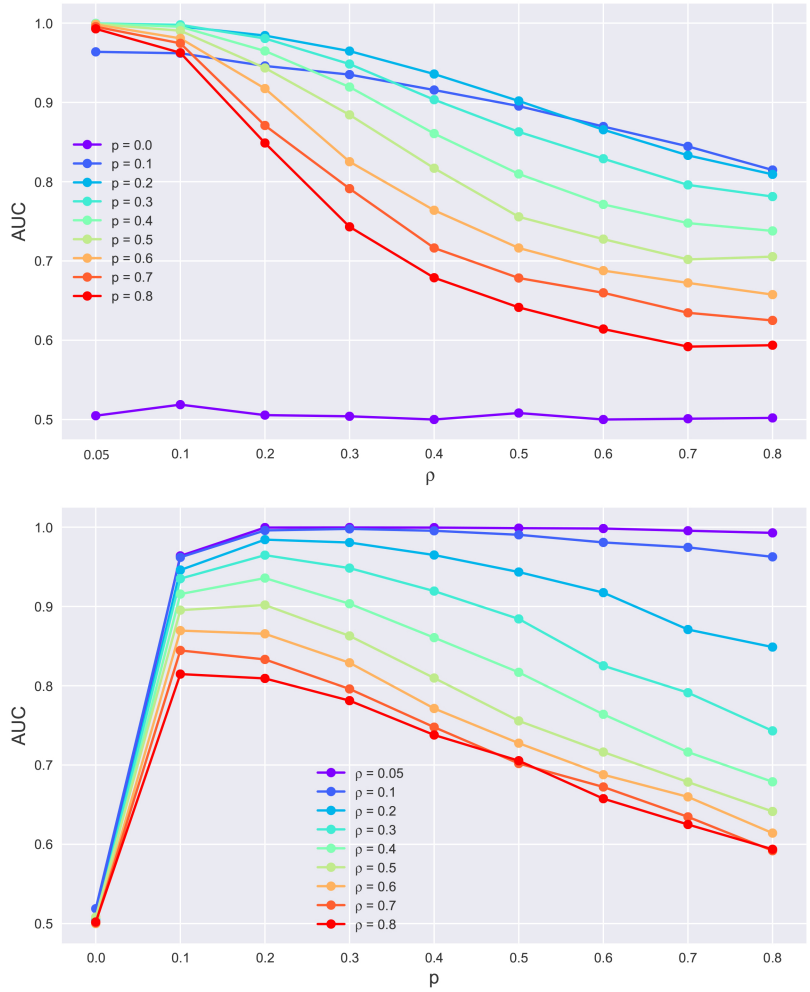

FIGURE 3.4.: Dependence of the area under the ROC curve (AUC) for ECA-based network inference for Erdös-Rényi random networks with $N = 20$ nodes on the link density ρ and propagation probability p.

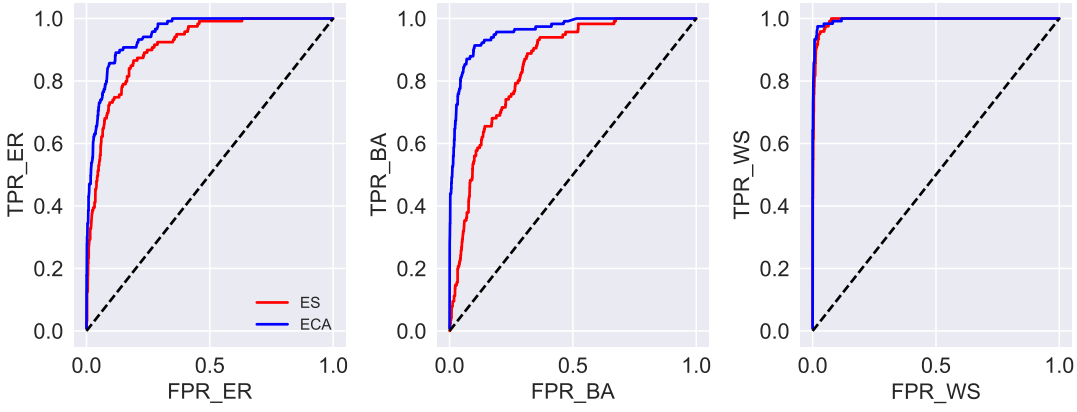

FIGURE 3.5.: ROC curves for network inference based on ECA (blue) and ES (red) for a single realization of the SIRS model on Erdös-Rényi, Barabási-Albert and Watts-Strogatz networks with $\pi = 0.3$ (from left to right) with $N = 60$ nodes and a mean node degree of 4 (i.e., a link density of $\rho \approx 0.068$).

$N = 30$ nodes) does not change the reported results qualitatively.

Further, to investigate the generality of our results obtained for the generic event propagation model, we perform the same type of analyses for simulations of the SIRS model (see Appendix A.1 for more details). Considering a microscopic version of the SIRS model, changes from one state to another depend on the individual's current state and contemporary interactions with their neighbors on the network. Susceptible individuals can become infected by an infected neighbor with a fixed probability β. Hence, the higher the number of infected neighbors at time t, the more probable a node becomes infective itself. By choosing the infectious period to be larger than the immune period (during which an individual can get reinfected by any of its neighbors with nonzero probability), we guarantee that the spreading process at the individual node's scale presents an oscillatory behavior. We note that under the considered parameter values, there have been some simulation runs leading to a saturated dynamics since the infectious state has completely disappeared from the network. The results presented in the following therefore refer to only a single realization where non-trivial dynamics has been observed over the whole simulation period. For ECA, we have chosen the parameters $\Delta T = 1$ and $\tau = 0$, reflecting that there is no intrinsic propagation delay as in the previously considered generic event propagation model. Fig. (3.5) demonstrates that under a parameter setting with a relatively small link density, network inference based on ECA outperforms the results based on ES for Erdös-Rényi and Barabási-Albert networks. In contrast, the results are almost equivalent for small-world networks generated via the Watts-Strogatz algorithm. This general tendency qualitatively agrees with the results of the generic event propagation model, though exhibiting certain differences in the details. Since the underlying network topologies are generally the same for both models (despite the different parameter values), any observable difference might point to effects due to the specific dynamics differing among the two considered models.

3.4. Summary

In the last couple of years, there have been considerable methodological advances in understanding how different topologies shape the resulting dynamics of different types of processes on a complex network. Among others, much work has been devoted to different types of spreading dynamics[166], like in the context of epidemiology, and cascading failures of functional units (e.g., in electricity or communication networks). Here, we addressed the inverse problem of network inference from observed event series reflecting such kind of spreading phenomena on an unknown network structure. Drawing upon the idea of functional connectivity serving as a proxy for structural connectivity, we presented an inter-comparison between two similarity measures for event sequences: event synchronization (ES) and event coincidence analysis rates (ECA). Both methods are based on the same general idea, while differing mainly in utilizing a static (ECA) versus data-adaptive "dynamic" coincidence interval (ES) to define simultaneity of events. Our results revealed that both ES and ECA, indeed, allow an accurate inference of the underlying network structures from the pattern of events without employing any prior knowledge on the type of the observed spreading dynamics. However, theoretical reasoning suggests that in specific situations with events arising as clusters in time (i.e., with very short inter-event waiting times) ES might underestimate the actual number of simultaneous events in two sequences. On the other hand, ECA does not suffer from the same issue as the fixed coincidence interval of ECA may still identify events with a certain mutual time lag as being statistically related.

4. Reconstructing dynamics of Complex System Using Stochastic Differential Models

Generally, real-world systems are subjected to noise and stochastic fluctuations (e.g., thermal fluctuations or internal dynamics) that profoundly affect the overall evolution of such systems. To consider the effect of these fluctuations on the underlying dynamics, stochastic differential equations (SDEs) have been employed for various random phenomena such as turbulent cascades [57], nano-friction fluctuations [93], and molecular dynamics [76]. In the framework of SDEs, one can solve the governing motion equation for a reduced-dimensional system by explicitly modeling macroscopic variables while microscopic variables are represented by noise. A well-known approach to retrieving such SDEs from small sets of observed time series is to reconstruct the drift and diffusion terms of a Langevin equation from the data-derived Kramers-Moyal (KM) coefficients. For systems where interactions between the observed and the unobserved variables are crucial, the Mori-Zwanzig formalism (MZ) allows deriving Generalized Langevin equations (GLEs) that contain non-Markovian terms representing these interactions. Similarly, the Empirical Model Reduction (EMR) approach has more recently been introduced. The results presented and the figures shown in this chapter are based on publication (Hassanibesheli, F., Boers, N., and Kurths, J. (2020). Reconstructing complex system dynamics from time series: a method comparison. New Journal of Physics, 22(7), 073053)

This chapter is organized as follows: In the first part, we systematically investigate the statistical properties of different dynamical systems using the three aforementioned methods (KM, MZ, and EMR) (see section 2.2 in chapter 2). To do so, we reconstruct the dynamical equations of motion of various synthetical and real-world processes obtained from these three approaches. In the second part, we utilize KM expansion to analyze abrupt transitions in paleoclimatic records that can be observed in δ^{18}O and dust counts in Greenland's ice sheet. Additionally, we investigate the possible coupling between the temperature and the concentration of dust in the atmosphere. The results presented in this chapter are based on publication (Rydin, L., Riechers, K., Hassanibesheli, F., Witthaut, D., Lind, P. G., and Boers,N. (2021). Data-driven Reconstruction of Last Glacials' Climate Dynamics Suggests Monostable Greenland Temperatures and a Bistable Northern Hemisphere Atmosphere. Earth Syst. Dynam. Discuss. in review.)

4.1. Results and Discussion Part I

4.1.1. Unimodal Synthetic Time Series

As a first example, we begin by considering the dynamics of the Ornstein–Uhlenbeck (OU) process, which is a stationary Gaussian-Markov process with continuous paths. Analogous to Brownian motion, which is the scaling limit of a random walk, the OU

process is known as the scaling limit of the Ehrenfest urn model [100] describing the diffusion process as a Markov chain. An OU process, satisfies the following SDE:

$$\frac{dx(t)}{dt} = \theta(\mu - x(t)) + \sigma\eta(t) \tag{4.1}$$

where $\eta(t)$ denotes a Gaussian white noise with $<\eta(t)\eta(t')> = \delta(tt')$. Here σ is the volatility of the process that controls the intensity of the noise. The OU process is known to be mean-reverting, where the drift coefficient controls the forcing of the state variable $x(t)$ back to its mean μ with the rate of mean reversion of θ. When approaches zero, the conventional Gaussian white noise is recovered. Since finding the analytical solution of stochastic differential equations is challenging, numerical solutions, such as Euler-Maruyama and the Euler-Milstein methods, can be considered an alternative to approximate the solution. In this study, to numerically integrate Eq. 4.1, we employ the Euler-Maruyama method that is based on stochastic calculus. Through stochastic integration, the discrete-time evolution of $x(t)$ can be written as follows:

$$x(t+1) = x(t) + \theta(\mu - x(t))\Delta t + \sigma\Delta\eta(t) \tag{4.2}$$

where $\Delta\eta(t)$ is independent identically distributed Wiener increments and can be approximated by$\sqrt{\delta}\mathcal{N}(0,1)$.

In addition, we consider an OU process subjected to multiplicative noise [84, 171, 202], where fluctuations are modeled as a function of a random variable. For a system subjected to an additive noise the deterministic potential corresponds to stochastic steady-state potentials through $V_s = \frac{V_d}{\sigma}$, where σ is the intensity of fluctuations that are independent of the system state x. It implies that the system exhibits stochastic fluctuations from its deterministic attractor in response to the stochastic force. In the case of multiplicative noise, a noise-induced drift becomes visible, and consequently, a new attractor basin is generated that does not exist in the absence of state-dependent fluctuations. If we substitute the additive noise in Eq. 4.1 with a multiplicative noise (i.e., the diffusion term is multiplied by $1 + x^2$) whose variance depends on the state variable $x(t)$:

$$\frac{dx(t)}{dt} = \theta(\mu - x(t)) + (1 + x^2)\eta(t) \tag{4.3}$$

We recall that for stochastic integration, there are two different approaches, Itô and Stratonovich (see Appendix B.1 for more details). In the case of the OU process, we observed that these two interpretations are equivalent. To generate an OU process, we produce 10^6 data points with $\theta = 1$, $\mu = 0$, $\sigma = 0.5$ and time sampling $\Delta t = 0.01$. Panels (e) and (f) in Fig. (4.1) illustrate the generated OU processes subjected to additive and multiplicative noise, respectively.

We begin by investigating the non-parametric estimates of the first $D_1(x)$ and second $D_2(x)$ KM coefficients (see Eq. 2.19). Panels (a) and (c) in Fig. (4.1) show the KM coefficients of an one-dimensional time series generated by Eq. 4.1. Panels (b) and (d) illustrate the $D_1(x)$ and the corresponding $D_2(x)$ of an OU process subjected to a multiplicative noise (see Eq. 4.3). As it can be seen, the drift and diffusion terms of both OU processes can be successfully restored. We usually do not have access to infinite time resolution. Hence, to investigate the effect of finite sampling interval on the estimation of $D_1(x)$ and $D_2(x)$, we compared the analytical solution of AFPE (see Appendix B.2) with the estimated KM coefficients for OU processes subjected to an additive and multiplicative noise based on Lade [114, 117] interpretation. As can be concluded from top (bottom) panels of Fig. (4.2), the values of the KM coefficients estimated from discrete times series of an OU process with additive (multiplicative) noise and the corresponding exact solution from AFPE are in a good agreement.

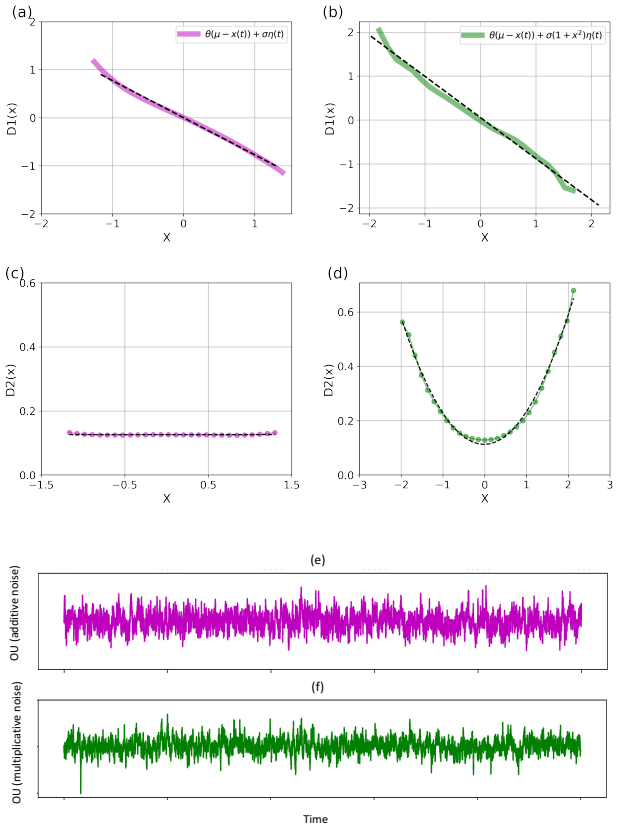

FIGURE 4.1.: The non-parametric estimates of KM coefficients. The left column illustrates the associated $D_1(x)$ and the second $D_2(x)$ KM coefficients for an OU process subject to arbitrary additive noise. Panels (b) and (c) in the right column show the corresponding estimated KM coefficients for an OU process subjected to the multiplicative noise of form $1 + x^2$. (e) and (f) display original time series of OU processes subjected to additive and multiplicative noise respectively.

Using the first and second KM coefficients, we can derive the LE (see Eq. 2.21) for the two OU processes mentioned above. Additionally, we reconstruct the time evolution of target systems' underlying dynamics using the GLE and EMR (for definition, see Eq. 2.32 and Eq. 2.37). We note that throughout this study, the deterministic terms of simulated time series constructed by the GLE and EMR are estimated directly from the first KM coefficient. Assuming that the deterministic function of the GLE, $\Omega(x_t)$, is known explicitly, we obtain the memory kernel $\mathcal{K}$ in Eq. 2.32 as follows:

$$C^k_{y,x} = -\sum_{l=1}^{k} \mathcal{K}^l C^{k-l}_{x,x}, \tag{4.4}$$

where $y_t = x_t - x_{t-1} - \Omega(x_{t-1})$. After estimating the memory kernel from the known correlation structure, we can easily calculate the noise term in Eq. 2.32 from the available time series.

Fig. (4.3) displays the statistical properties of the time series that obey the OU dynamics obtained from the three different inverse modeling methods (KM, MZ, and EMR) in comparison with the original time series. Here, we obtain the PDFs as averages over 1000 simulated time series, and the error bars are based on the $\pm\sigma$ deviations

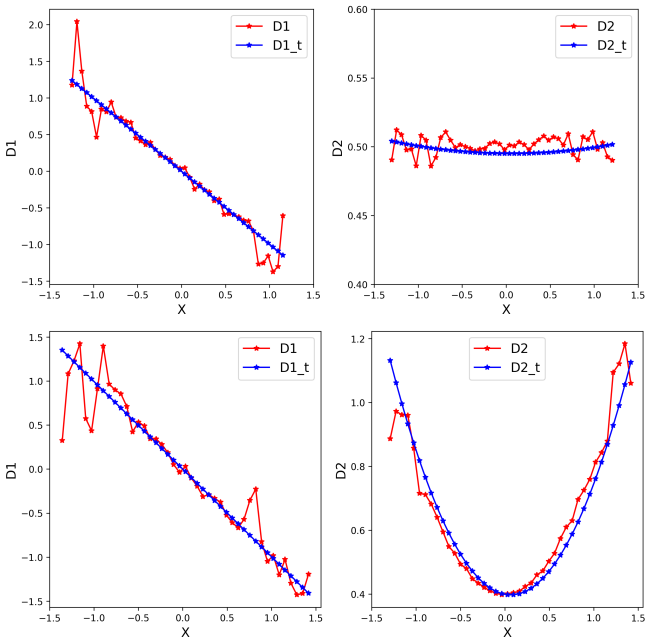

FIGURE 4.2.: Comparison of the estimated drift ($D_1(x)$) and diffusion ($D_2(x)$) coefficients with corresponding exact solutions for an OU process containing 10^6 points with sampling frequency of 10^{-2}. (a) Estimated $D_1(x)$ and $D_2(x)$ (red) and the values obtained from the analytical solution via AFPE (blue) for an OU process with additive noise. (b) Same as (a) for an OU process subject to the multiplicative noise $1 + x^2$.

around the averages for these 1000 realizations. According to Fig. (4.3) all three SDE models could perfectly reproduce samples of series obeying the statistic of the underlying dynamics. Here the auto-correlation exponentially decays, and the memory coefficients of GLE are zero except at $\tau = 1$, which indicates the process is Markovian. Similarly, we repeat the analysis for an OU process subjected to multiplicative noise. As it can be observed from Fig. (4.4), all three methods work very well in reproducing the statistical features of the linear system also when exposed to symmetric multiplicative noise.

For more comparison, we analyzed a system subjected to asymmetric noise, which results in a skewed distribution. We thus multiply the noise term in Eq. 4.1 by $(1+x)$ (instead of $(1+x^2)$) and evaluate the statistical properties of the system. According to Fig. (4.5), the KM model exhibits better performance in comparison to the two other models since it can directly estimate the (in this case, asymmetric) state-dependent noise from the time series.

So far, we have considered additive and multiplicative noise, with η given by Gaussian white noise. However, in many physical and biological systems, fluctuations exhibit some degree of correlation that cannot be explained by uncorrelated white noise. Therefore we substitute the stochastic term of Eq. 4.1 with a first-order autoregressive process AR(1) and investigate the performance of the three methods mentioned above to derive SDEs in the presence of colored noise. An AR(1) process is given by:

$$Y_{t+1} = \alpha Y_t + \eta(t) \tag{4.5}$$

where η is a Gaussian white noise process with zero mean and constant variance. According to Fig. (4.6), the statistical properties of simulated time series constructed via the KM method can not perfectly follow the original ones. From Tables B.1 and B.2 it can also be concluded that MZ and EMR perform better (with smaller MSEs).

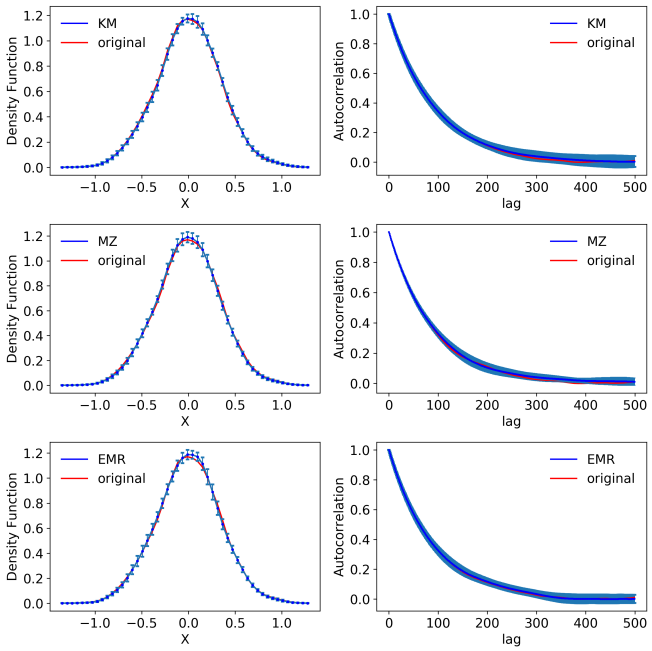

FIGURE 4.3.: A stochastic process with linear drift and additive noise. (a) Original (red) and randomly chosen simulated time series based on KM, MZ, and EMR methods (from top to bottom, respectively). (b) Summary statistics (PDFs in the left column and ACFs in the right column) of original and simulated time series derived from 1000 sample time series reconstructed by the three stochastic models (KM, MZ, and EMR), from top to bottom as indicated in the legend. The original system is an Ornstein–Uhlenbeck process with $\Delta t = 0.01$ whose statistical features are shown in red color.

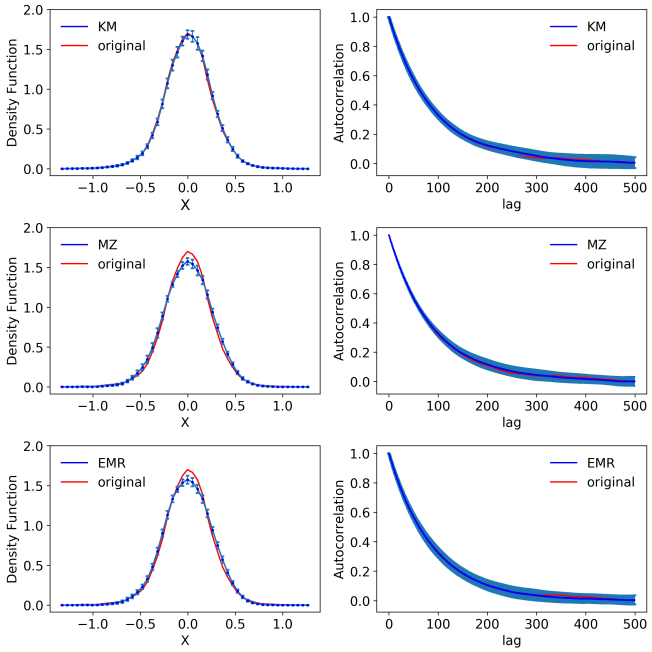

FIGURE 4.4.: Statistical properties of the observed and simulated OU with linear drift and multiplicative noise $(1 + x^2)$. The left column illustrates the PDFs for simulated results obtained as averages over 1000 realization with uncertainties in blue color. The right column displays the ACFs of ensembles of simulated time series constructed based on the three different stochastic models (KM, MZ, and EMR, from top to bottom).

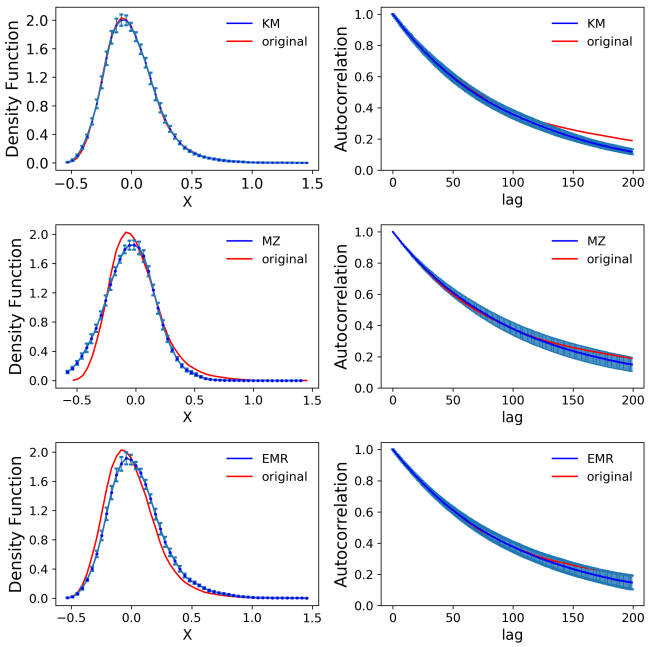

FIGURE 4.5.: Statistical properties of the observed and simulated OU process subject to the asymmetric noise $(1 + x)$. The left column illustrates the PDFs for simulated results obtained as averages over 1000 realization with uncertainties in blue color. The right column displays the ACFs of ensembles of simulated time series constructed based on the three different stochastic models (KM, MZ, and EMR, from top to bottom).

This is expected, since the presence of colored noise implies a deviation from the white-noise assumption of the LE and the KM method to derive it.

It is also of interest to evaluate the performance of these SDEs when the system exhibits some short-term memory in the deterministic part. For this purpose, we consider a stochastic delay differential equation given by:

$$\frac{dx(t)}{dt} = \alpha x(t) + \beta x(t - \tau) + R(t) \tag{4.6}$$

The random force R is given by serially correlated noise produced by a second-order autoregressive AR(2) process in which the current value depends on the two previous values. As can be observed from Fig. (4.7), the EMR and MZ methods yield significantly better approximations of the PDF and ACF than the KM method for this system with short-term memory in the drift and colored noise. This provides an instructive example for situations where the MZ and EMR approaches clearly outperform the KM approach (as theoretically expected).

4.1.2. Bimodal Synthetic Time series

Until now, we studied processes with a unique mode; however, there are various stochastic dynamics in natural systems that do not exhibit uni-modality. We start with generating a synthetic process in which a particle moves in a double-well (DW) potential at $\pm\sqrt{\theta}$, driven by additive Gaussian noise.

$$\frac{dX(t)}{dt} = \theta(x(t) - x^3(t)) + \sigma\eta(t) \tag{4.7}$$

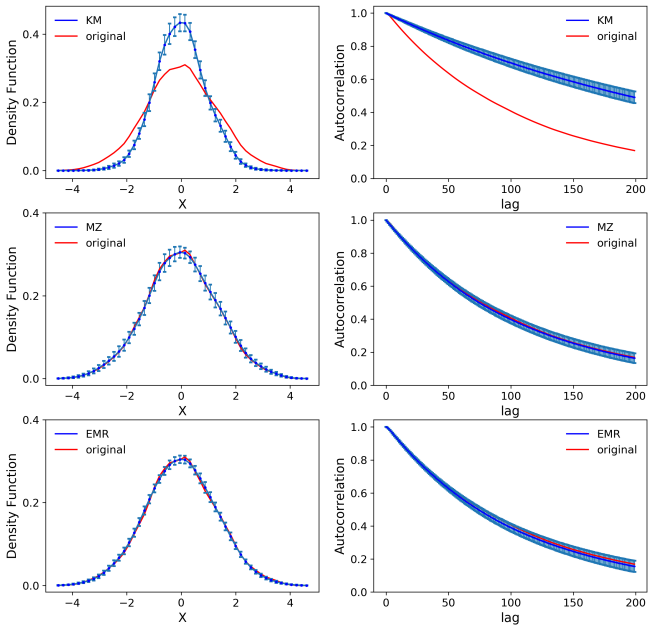

FIGURE 4.6.: Summary statistics of an OU process with linear drift and colored noise time series and simulated time series reconstructed by the three stochastic models (KM, MZ, and EMR) from top to bottom as indicated in the legend.

According to Eq. 4.7, we generated a process consists of 10^5 points with sampling interval $\Delta t = 0.01$. The summary of non-parametric estimations of the first and second KM coefficients of the process can be found in Fig. (4.8a) and Fig. (4.8c), respectively. Considering the Markov property of the underlying system, the KM analysis captures precisely the nonlinear interactions taking place in the macroscopic scale (see Fig. (4.8a)). In Fig. (4.9), we evaluated the performance of the three SDEs mentioned above in reconstructing the dynamics of the underlying system. The left column in Fig. (4.9) illustrates the average PDFs of model-simulated time series in comparison with the original time series for 1000 realizations, and the right column displays the associated ACFs. It can be seen that these three stochastic models can mimic the key features of the original time series very well.

We repeated the KM analysis for a DW process subjected to a multiplicative noise, see Fig. (4.8b) and Fig. (4.8d). In the presence of a multiplicative noise, estimating the underlying dynamic is not trivial anymore for the EMR and GLE. As it can be observed in Fig. (4.10), the average ACFs deviate from the original ones for the MZ and EMR methods, while the KM method still achieves an excellent approximation of the PDF and ACF.

4.1.3. Real-World systems

To complement our analysis, we continue by studying three real-world time series, i.e., $S\&P500$ stock index, El Niño-Southern Oscillation (ENSO), and the concentration of $[Ca^{2+}]$ in the Greenland's ice core.

4.1.3.1. $S\&P500$ Stock Index

In recent years, quantifying stochastic dynamics of financial time series (e.g., stock prices and stock market indices) using SDEs to describe their evolution has attracted considerable attention[52, 61]. In this study, we analyzed the weekly $S\&P500$ stock

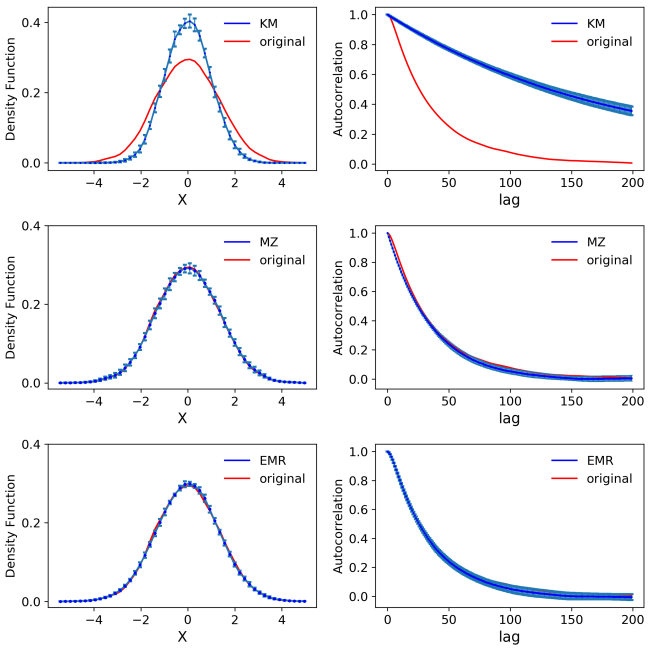

FIGURE 4.7.: Statistical properties of the observed (red) and simulated (blue) time series from a delayed OU process, with $tau = 2$, subject to correlated noise produced by an AR(2) process. The simulated time series were constructed using KM, MZ, and EMR methods (from top to bottom), averaged over 1000 different realizations.

index for the time span of 35 years (1950-1985) and analyze its distinctive statistical properties. Here the stock return price $ln\frac{p(t+\delta t)}{p(t)}$ represents the state variable $x(t)$. The statistical results of the simulated time series estimated by the three inverse modeling methods are shown in Fig. (4.11). According to the KM coefficients, the deterministic part of the dynamics is described by a linear function of the state variable, while the stochastic term exhibits nonlinear behavior.

In this first real-world case, the approximation of the PDFs is not perfect anymore. The results reveal that the MZ formalism outperforms the two other methods in terms of the ACF, with the perfect agreement to the original return series (see Fig. (4.11), second row, second column). It is worth mentioning that the statistical properties of financial time series exhibit a time-scale dependence and long-range correlations [63, 115]. Therefore we conclude that the memory effects of recent returns occurring in different time scales (from minutes to several days) can be modeled well by the MZ technique.

4.1.3.2. El Niño-Southern Oscillation (ENSO)

Another empirical time series that we use is the Niño-3 index [180], which is one of several ENSO indicators of tropical Pacific sea surface temperatures (SST). During the last decades, understanding the mechanisms underlying ENSO variability and prediction of future fluctuations has attracted substantial attention [2, 30, 94, 108]. ENSO describes variations in temperature and pressure in the eastern Pacific ocean and has significant impacts on global climate variability. We reconstruct the Niño-3 monthly sea surface temperature (SST) index averages across (5°N–5°S, 150°–90°W) from 1891 to 2015 using the KM, MZ, and EMR inverse modeling approaches. The Non-Gaussian behavior of Niño-3 indicates a nonlinear process, quantified by the positive skewness of the SST distribution, which may reside in the interaction of oceanic variables of

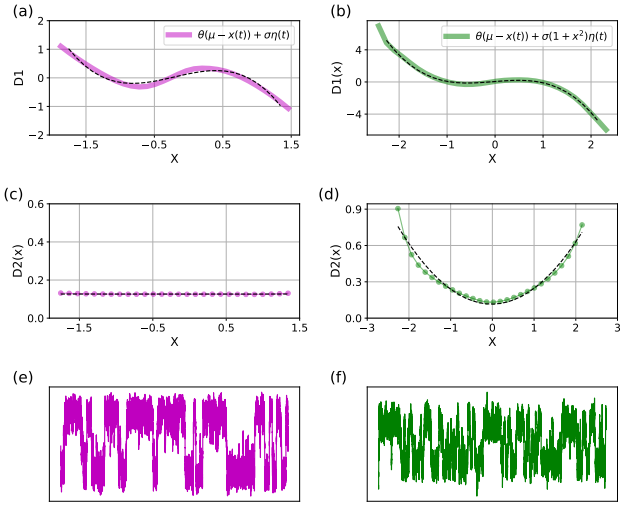

FIGURE 4.8.: The non-parametric estimates of KM coefficients. The left column illustrates the associated $D_1(x)$ and the second $D_2(x)$ KM coefficients for a process in a double-well potential subject to arbitrary additive noise. Panels (b) and (c) in the right column show the corresponding estimated KM coefficients for a process in a double-well potential subject to the multiplicative noise of form $1 + x^2$. (e) and (f) display original time series of processes in a double-well potential subject to additive and multiplicative noise respectively.

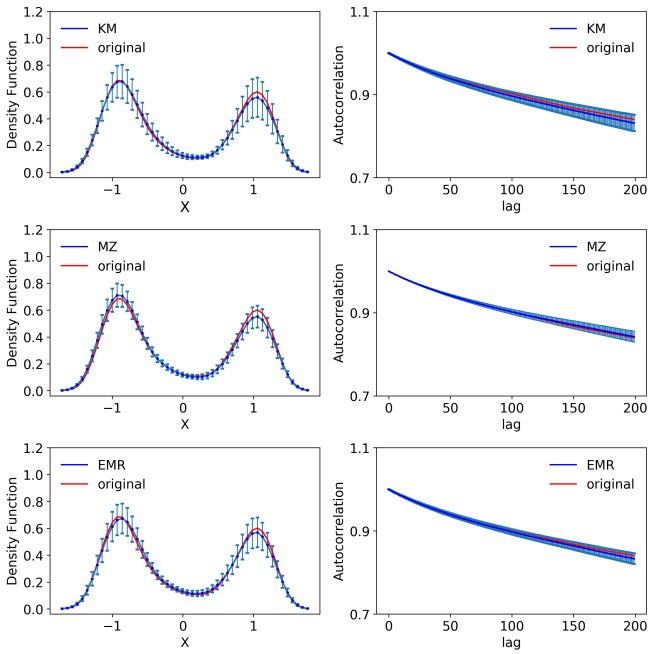

FIGURE 4.9.: Comparison of the numerical simulation of a process in a double-well potential subject to an additive Gaussian noise with intensity $\sigma = 0.5$,10^6 data points and a sampling interval of 0.01. The plot displays PDFs and ACFs of the original(red) and simulated(blue) time series obtained from three different stochastic models (KM, MZ, and EMR as indicated in the legend), averaged over 1000 realizations.

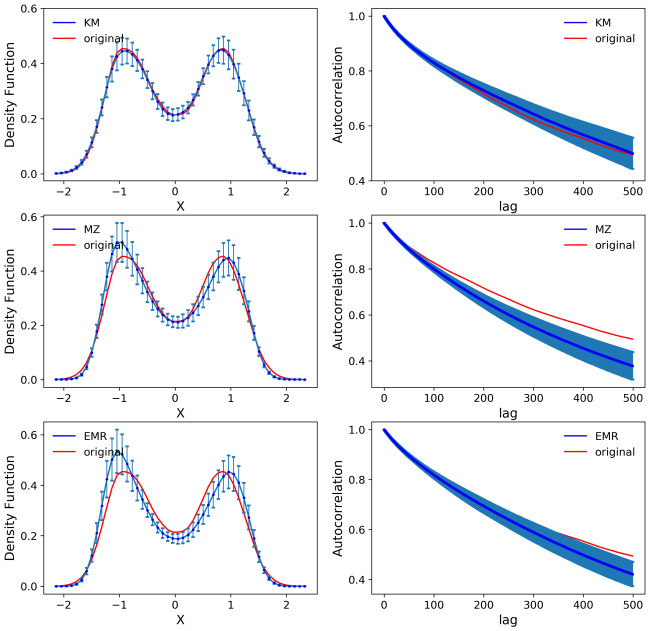

FIGURE 4.10.: Statistical properties of the observed (red) and simulated (blue) time series related to a particle motion in a double-well potential, subject to multiplicative noise. The simulated time series were constructed using KM, MZ, and EMR methods (from top to bottom), averaged over 1000 different realizations.

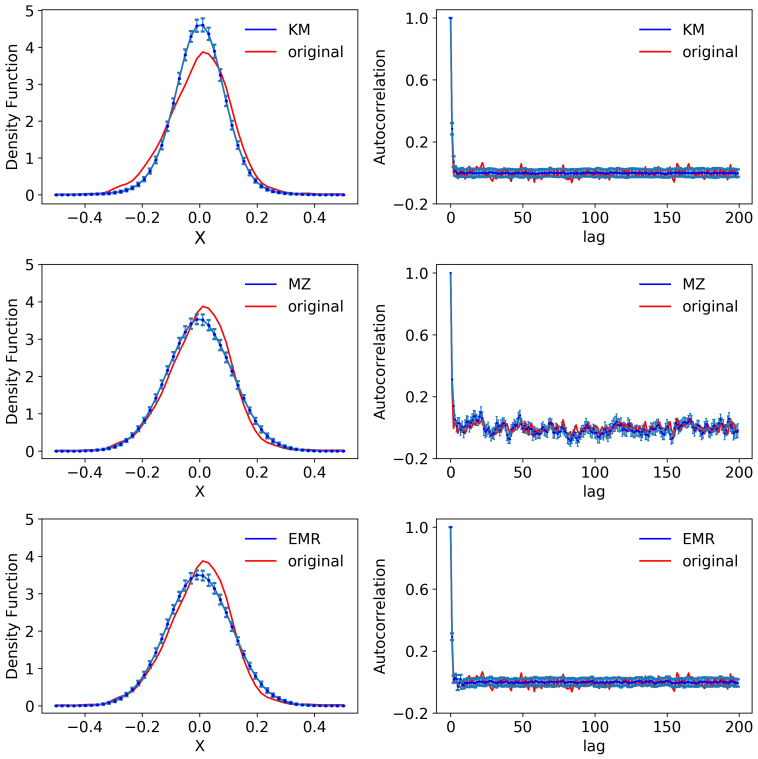

FIGURE 4.11.: Summary statistics of $S\&P500$ stock index and simulated time series over 1000 realizations (blue) created by KM, MZ, and EMR models from left to right, respectively. The statistics plotted in red correspond to the original weekly $S\&P500$ stocks index for the time span of 35 years (1950-1985).

interest and the fast atmospheric forcing [144]. This interaction introduces memory into the system dynamics that should not be expected to be fully captured by models established based on Markovian assumptions. Fig. (4.12) displays the PDFs and ACFs of the resulting simulated time series. Although all three inverse methods produce similarly skewed and heavy-tailed distributions, comparing the AFCs, MZ achieves a higher accuracy approximation than KM and EMR. We repeated this procedure for the Niño-4 index and also, in that case, found that the MZ approach outperforms the other two methods. According to these results, the most significant characteristic of these real-world processes leading to ENSO variability is the presence of serial correlations connected to internal interactions between observed (slow) and unobserved (fast) variables, which can be captured best by the MZ approach.

To investigate whether our results are prone to overfitting, we first calibrated the three SDEs on the first half of the time series. Then, to validate the parameters, we compared the resulting time series statistics with the ones of the second half of the time series Niño-3, see Fig. (B.2). It can be inferred that the longer-term variations in the ACF are still captured to some degree by the MZ method. For the ENSO case, the slow variations of the ACF correspond to low-frequency variability present in this time series, which is not noise but an essential part of (internal) climate variability. The fact that the reproduction of these slow variability modes is less accurate when only calibrating on the first half of the time series is due to the fact that capturing these slow variability modes is harder when considering only a shorter part of the time series.

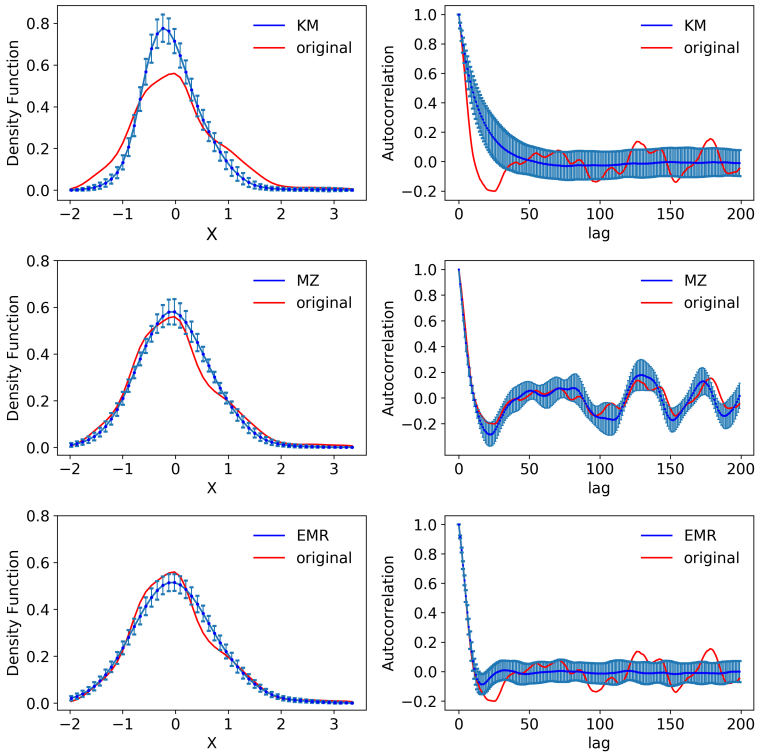

FIGURE 4.12.: Comparison of model performance of KM, MZ, and EMR in reconstructing dynamics of Niño-3 monthly sea surface temperature (SST) index. The figure depicts the statistical properties of the observed (red) and simulated (blue) monthly Niño-3 SST indices from 1891 to 2015.

4.1.3.3. Ca^{2+} Proxy

As the final example, we investigated the high-resolution (20yr-average) Ca^{2+} (interpreted as a proxy for atmospheric circulation patterns), collected from the NGRIP ice core on the GICC05 time scale [177]. Because of the substantially better signal-to-noise ratio, we focused here on the Ca^{2+} time series between 60ka and 30ka b2k. We apply the KM, MZ, and EMR approaches to reproduce dynamical and statistical properties of the underlying dynamics of Ca^{2+}. Fig. (4.13) shows the statistical properties (PDFs and ACFs) of the observed and simulated Ca^{2+} time series. Because of the large amplitude of the Ca^{2+} concentration variations, the calculations were conducted in natural logarithmic scale [19]. The results presented in Fig. (4.13) show that the modeled time series for all three methods could accurately reproduce the bimodality of the observed PDFs. However, it should be noted that the LE derived with the KM approach approximates the observed PDF better than the GLE and EMR approaches. The depth of the two potential wells in MZ and EMR are shallower than the observed PDF suggests. It is clear from the results illustrated in the right column of Fig. (4.13) that the MZ and EMR methods have better performances to construct the underlying auto-correlation structure of the Ca^{2+}. To study the underlying dynamics of Ca^{2+} time series we employed Stratonovich calculus instead of Itô (see Appendix B.1.1). In the Stratonovich calculus, white noise is approximated by continuously fluctuating noise with finite memory, which may be more suitable for approximating real-world time series [1].

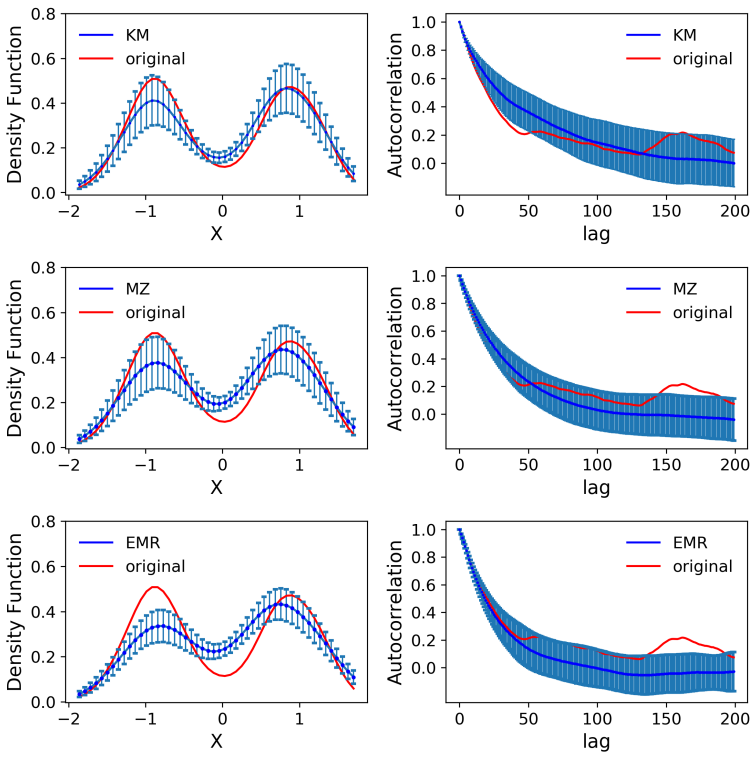

FIGURE 4.13.: Statistical properties of the observed and simulated Ca^{2+} time series in the interval between 60ka and 30ka b2k. The PDF of the original data is shown in red color while simulated time series obtained from KM, MZ, and EMR models (from top to bottom) can be found in blue color. The PDFs for the three stochastic models are averaged over 1000 realization and therefore considerably smoothed.

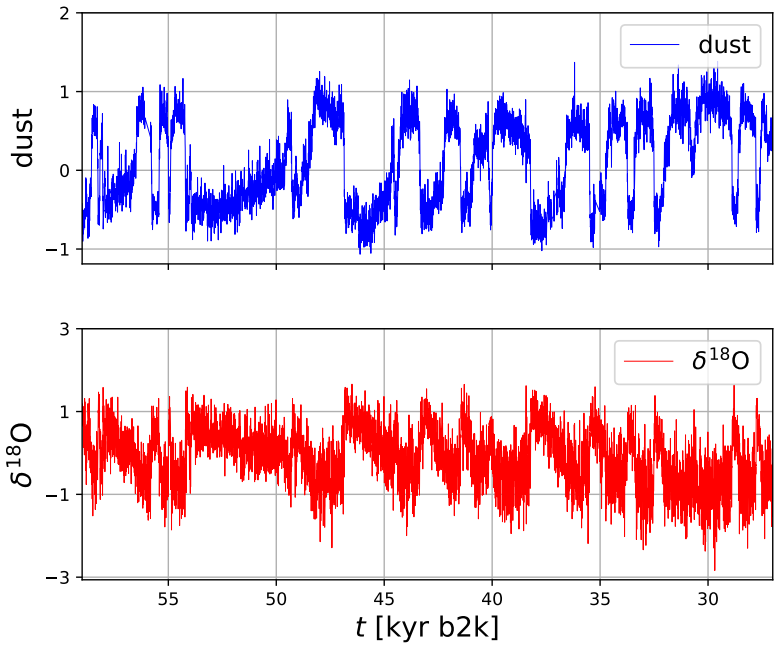

FIGURE 4.14.: The dust (blue) and the δ^{18}O (red) records from the NGRIP ice core in Greenland. The dust data is the natural logarithm of the actual dust concentrations.

4.2. Results and Discussion Part II

4.2.1. Dansgaard–Oeschger Events (DO)

Understanding the triggering mechanism of sudden climate transitions in the past is essential to better identify abrupt warming events and sea ice loss in the future. One of the best-documented examples is abrupt temperature shifts during the last glacial period (roughly from 115.000 years to 12.000 years BP), known as Dansgaard-Oeschger (DO) events which involve a wide range of time scales [19, 20, 41, 43, 113, 126, 127]. These rapid and strong warming events (ranging from 8 to 16.5°) followed by slow-paced cooling phases can be inferred from the content of stable isotope composition of water $\delta^{18}O$ in different Greenland ice cores. The concentration of $\delta^{18}O$ in the ice core provides reliable information about the evolution of Arctic temperature. Other properties of the ice, such as dust content and concentration of Calcium ions, can also be used to study these swift changes [143] (See Fig. (4.14)). The climate of the last glacial period witnessed about 26 distinct DO events. It has been suggested that the time interval between two consecutive DO events is roughly about 1500 yr and seems to be sampled from a Poisson process [42]. It should be emphasized that the corresponding exponential distribution of waiting times does not support the existence of the periodic pattern. Many physical mechanisms have been proposed as candidates for explaining such abrupt transitions; however, the origin and leading causes of DO events are still debated. One classical hypothesis attributes the DO cycles to the swing between two stable states of Atlantic meridional overturning circulation (AMOC) [35, 175] due to freshwater perturbations. In this context, it has been shown that DO events could be triggered by freshwater from the melting of iceberg discharges [206]. Even though the AMOC hypothesis mostly reproduces the spatial patterns of the abrupt changes, due to the insufficient paleoceanographic data, the exact causal relation between DO events and changes in AMOC still remains unknown. In recenter years, alternative

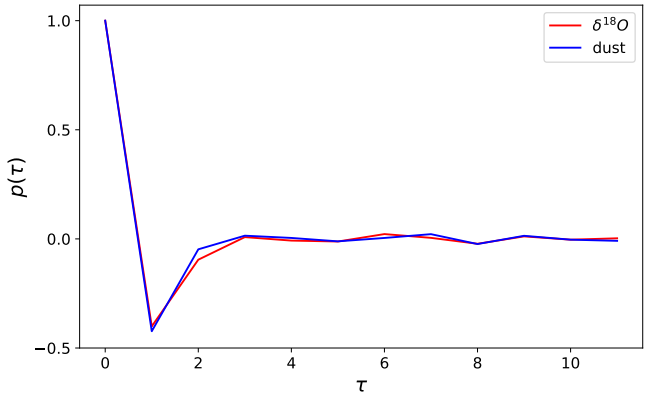

FIGURE 4.15.: Auto-correlation of the increments of δ^{18}O and dust records.

mechanisms such as sea-ice-shelf fluctuation [172], interhemispheric interactions [9], and internal variabilties [42] have also been introduced to generate DO events. In the context of dynamical systems, these abrupt changes may be induced by bifurcation or stochastic perturbations when a stable equilibrium is lost and a system tips to another attractor. Several studies have been conducted to simulate time series that reproduce the statistical and dynamical properties of paleoclimate records using data-driven stochastic models [19, 107]. In the following, we estimate the parameters of the stochastic process (i.e., drift and the stochastic diffusion components) driving the $\delta^{18}O$ and dust records using non-parametric KM analysis.

4.2.1.1. δ^{18}O and Dust Proxies in a One-dimensional Setting

We considered the paleoclimate records as the realization of a Markovian and stationary stochastic process [107, 108]. The ice measurements were taken at a fixed 5 cm and are not temporally uniform. Therefore, we interpolated the data to an equidistant time axis of 5 year intervals and fill the missing data with a next-neighbor interpolation (See Fig. (4.14)). Here, to evaluate whether the data of interest exhibits Markov property, we investigated the behavior of the auto-covariance function of its increments. To do so, we calculated the differences $\Delta x_t = x_{t+1} - x_t$, and obtain the auto-correlation function $\rho(\tau)$ as follows:

$$\rho(\tau) = \frac{\mathrm{E}\left[(\Delta x_t - \mu_t)(\Delta x_{t+\tau} - \mu_{t+\tau})\right]}{\sigma_t \sigma_{t+\tau}}, \tag{4.8}$$

As it can be concluded from Fig. (4.15), both δ^{18}O and dust records display weak anti-correlation at the lag $\tau = 1$, and no correlation for $\tau > 1$.

After fulfilling the Markovian assumption, we estimated the first and second KM coefficients of the dust and the δ^{18}O records derived from Eq. 2.17 in a one-dimensional setting. Panel (a) and (b) in Fig. (4.16) illustrate the estimated first KM coefficient for the dust and δ^{18}O respectively. As it can be observed, the drift term for the dust record exhibits a nonlinear behavior with three fixed points, while in the δ^{18} case, the drift has only a single stable fixed point. Panels (c) and (d) display the corresponding second KM coefficients of the dust and the the δ^{18} records.

To further understand the physical interpretation of the first KM coefficient, we took the integral over $D_1(x)$, which can be associated with the potential well where the drift

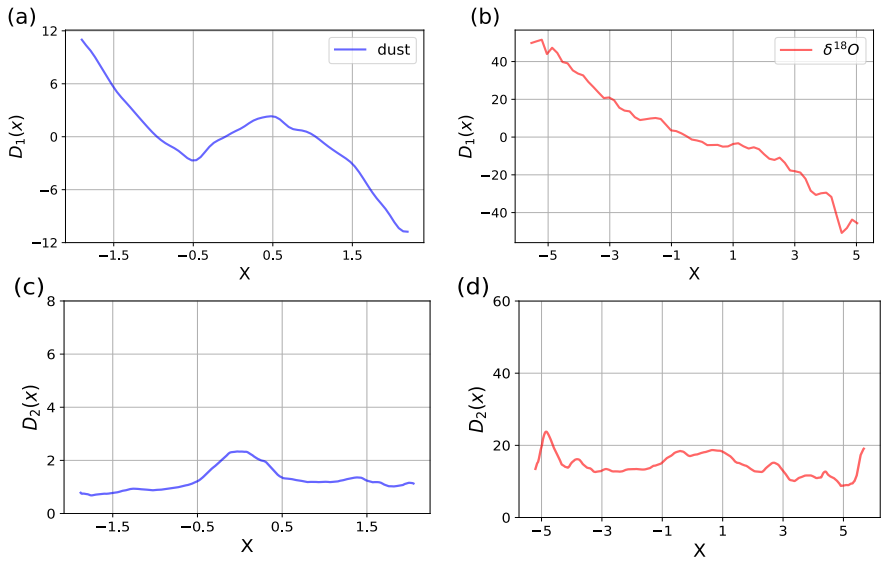

FIGURE 4.16.: The non-parametric estimates of the first and second KM coefficients for the δ^{18}O (panels a,c) and the dust (panels b,d) records in a one-dimensional setting.

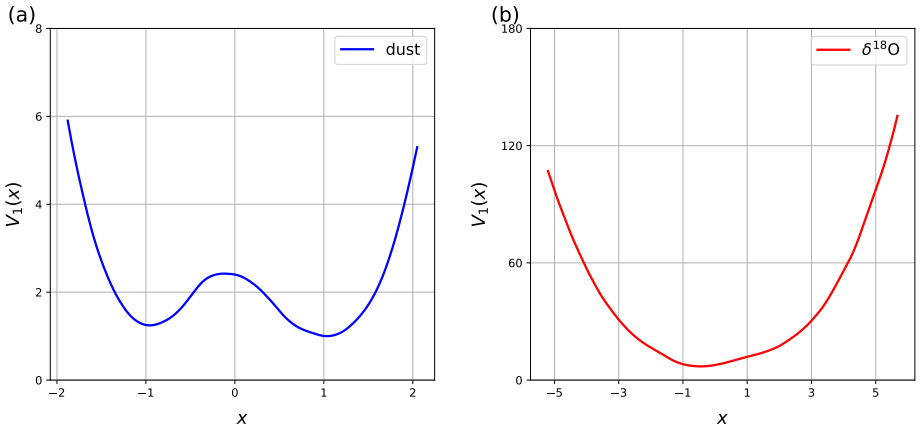

FIGURE 4.17.: Potential landscape of the drift for the dust and δ^{18}O records.

term drives a particle to the bottom of it

$$V(x) = -\int_{-\infty}^{x} D_1(x')\ \mathrm{d}x' + c. \tag{4.9}$$

We employed the concept of potential to show the stability configuration of δ^{18}O and the dust recordings. Fig. (4.17) displays the reconstructed potential wells associated with the δ^{18}O and the dust recordings. The suggested bistable dynamics for the dust recording can also be confirmed from the constructed potential with two minima shown in Fig. (4.17a). Here, the stochastic noise can trigger the switch from one stable state to another. Despite the apparent sudden regime shifts in the δ^{18}O records, the reconstructed potential well does not show bistability. According to Fig. (4.17b) δ^{18}O lives in a single-well potential with a fixed point around zero.

In the following, to distinguish continuous from discontinuous processes, we conducted two distinct analyses. According to the Pawula's theorem, we consider a

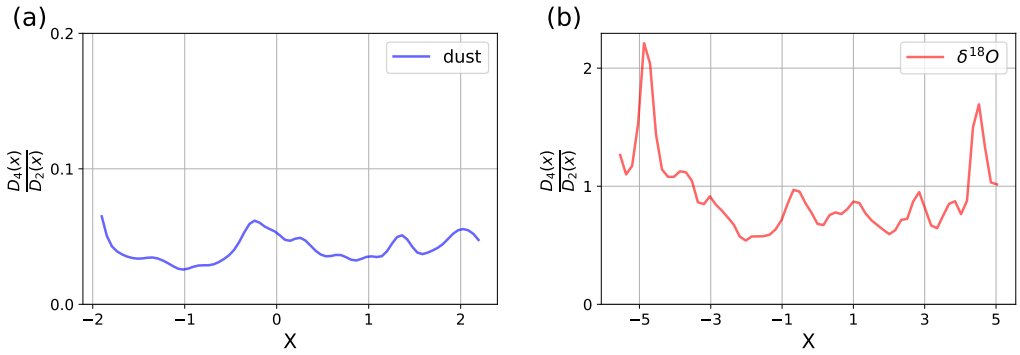

FIGURE 4.18.: The ratio $D_4(x)/D_2(x)$ for the dust (in blue) and the δ^{18}O (in red) records in a one-dimensional setting.

stochastic process $x(t)$ continuous if higher-order of KM coefficients ($D_m(x) = 0; m > 2$) are negligible. However, in various real-world systems, we observed non-vanishing higher-order KM coefficients. To investigate whether our data belong to the class of continuous processes or not, we calculated the ratio between the fourth and second KM coefficients ($D_4(x)/D_2(x)$) which explains the ratio of jumps to diffusive motion. As illustrated in Fig. (4.18a), the ($D_4(x)/D_2(x)$) ratio for the dust record is negligible, indicating the underlying dynamics can be regarded as a continuous stochastic process at the time scale of 5 yr. In other words, the time evolution of the dust record can be described by a Langevin equation in which the regime shifts are induced by stochastic force. On the other hand, for the δ^{18}O record, the non-negligible ($D_4(x)/D_2(x)$) ratio reflects the contribution of discontinuous events or jumps (see Fig. (4.18b)). The persistence fourth KM coefficient in the δ^{18}O may suggest a source of forcing other than Gaussian white noise, which could indicate an external trigger affecting Greenland temperatures directly. We note here that the interpretation of the fourth KM coefficient is not straightforward and Further analysis is required to determine the role of discontinuities in the δ^{18}O record.

To support our assessment regarding the continuity or discontinuity of a process, we calculate the Lehnertz–Tabar Q-ratio [114].

$$Q(x,\tau) = \frac{M_6(x,\tau)}{5M_4(x,\tau)} \sim \begin{cases} \tau, & \text{for diffusions,} \\ c, & \text{for jumpy processes.} \end{cases} \tag{4.10}$$

The Q-ratio allows distinguishing diffusive and jumpy behavior in time series in terms of the convergence of the conditional moments with the scaling τ. For a process with a discontinuous trajectory, $Q(x,\tau)$ exhibits no significant scaling relation with time t. In contrast, $Q(x,\tau)$ is linearly dependent on τ for a continuous process. In Fig. (4.19) we calculated the Q-ratio for the dust and the δ^{18}O records in a double-logarithmic scale. Accordingly, it can be seen that there is a linear relation between Q-ratio and time for the dust record while this quantity remains constant for δ^{18}O. Therefore, we can rule out that the underlying process of the δ^{18}O is diffusive.

4.2.1.2. δ^{18}O and Dust Proxies in a Two-dimensional Setting

As stated above, we used the KM analysis for the δ^{18}O and the dust records in a one-dimensional setting. Even though the trajectories of the dust and the δ^{18}O are very similar, we observed that the reconstructed potential of δ^{18}O does not explain regime shifts of the underlying dynamics. Two plausible scenarios can be considered leading to these abrupt transitions; 1) It can be due to the presence of correlated noise or 2)

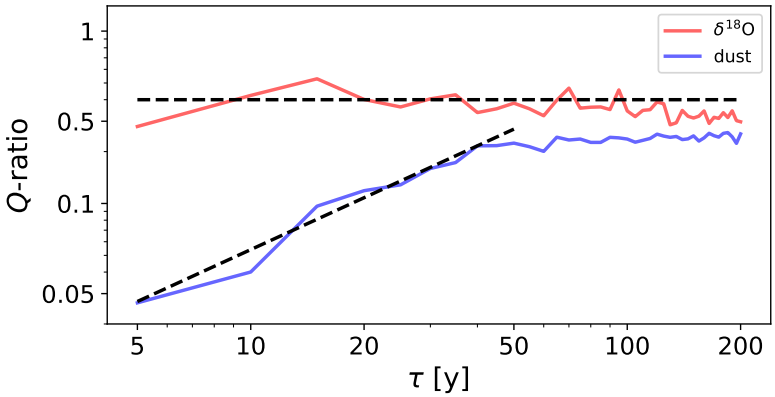

FIGURE 4.19.: The Q-ratio for the dust (in blue) and the δ^{18}O (in red) records.

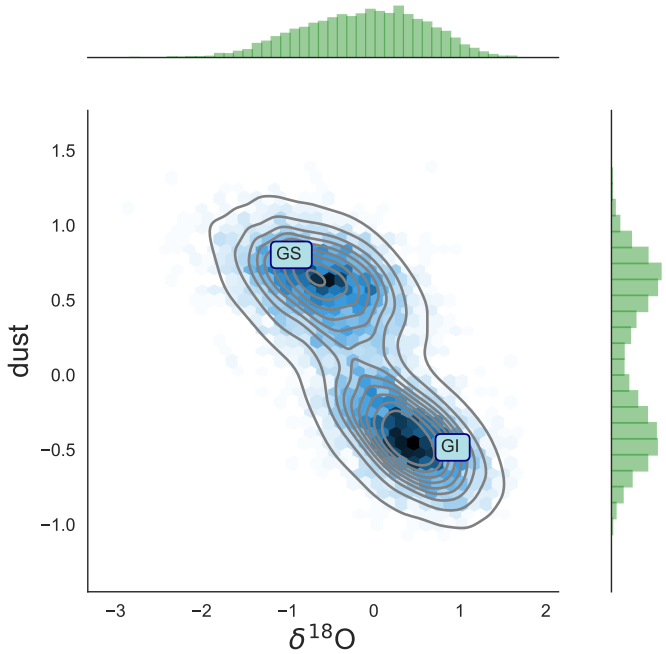

FIGURE 4.20.: Two-dimensional PDF of the δ^{18}O and dust records. The dotted elements are the records, separated into stadials (GS) and interstadials (GI).

possible coupling of dust and the δ^{18}O in which one record acts as an external control parameter for another one. In the following we employ a two-dimensional analysis (introduced in section 2.2.1.1) to investigate the coupling between these two proxies (see Fig. (4.20)).

Similar to the one-dimensional setting, we inspect the conditional potentials from the drift coefficients defined by:

$$V_{1,0}(x_1|x_2) = -\int_{-\infty}^{x_1} D_{1,0}(x_1', x_2)\, \mathrm{d}x_1' \tag{4.11a}$$

$$V_{0,1}(x_2|x_1) = -\int_{-\infty}^{x_2} D_{0,1}(x_1, x_2')\, \mathrm{d}x_2' \tag{4.11b}$$

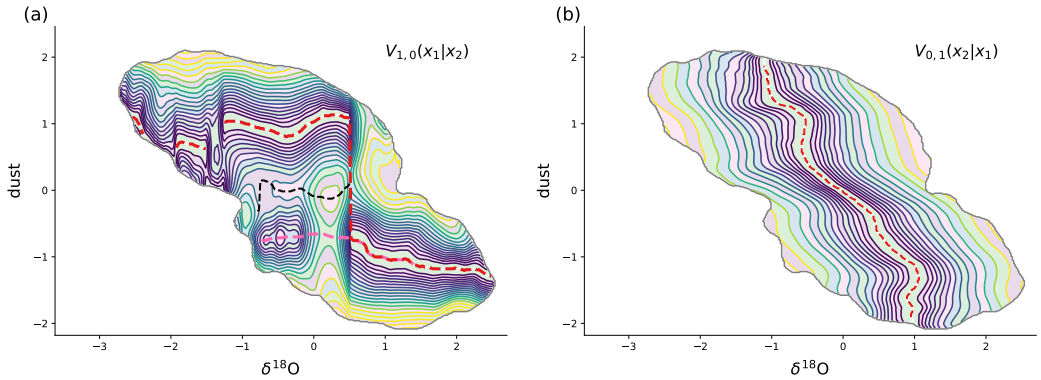

FIGURE 4.21.: In panel (a) the two-dimensional potential landscape $V_{1,0}(x_1|x_2)$ of the dust, conditional to the value of δ^{18}O has been shown. Depending on the value of δ^{18}O dust can dust can have three different fixed points. In panel (b) we illustrate the two-dimensional potential landscape $V_{0,1}(x_2|x_1)$ of the δ^{18}O conditioned on dust record .

Where $(V_{0,1})$ and $(V_{1,0})$ are the conditional potentials describing the motion of one variable conditioned on a fixed second dynamical variable. Fig. (4.21a) displays the reconstructed conditional potential $V_{1,0}(x_1, x_2)$ for the dust conditioned on δ^{18}O. The dust appears to have a different number of fixed points, changing from a mono-stable to a bistable regime, depending on the value of δ^{18}O. According to Fig. (4.21a), the type of regime-switching in the dust record, from stadial to interstadial phases, can be related to a double-fold bifurcation. Where for approximately $\delta^{18}\text{O} < -1.0$ there is only one stable fixed point (a global minimum), and for approximately $-1.0 < \delta^{18}\text{O} < 0.6$ there are three fixed points, two stable ones (a local minimum and a global minimum) and an unstable one (the local maximum) between them. In a similar manner we investigated the conditional potential $V_{1,0}(x_1, x_2)$ of the δ^{18}O conditioned to a value of dust (see Fig. (4.21b)). We found that the δ^{18}O has one minimum – given any value of the dust variable– indicating the characteristic of a mono-stable process. Notably, the position of the minimum of the δ^{18}O potential landscape is explicitly dependent on the value of the dust. Our results reveal that the abrupt changes are not intrinsic features of the δ^{18}O proxy and may stem from the coupling to other climate variables.

4.3. Summary

Stochastic differential equations (SDEs) are a promising approach for studying dynamics of complex systems in situations where only a few variables are actually measured. For some typical example settings, we studied the performance of three methods to derive SDEs – the Kramer-Moyal approach to derive Langevin equations (LE) with potentially multiplicative noise, the Mori-Zwanzig approach to derive Generalized Langevin Equation (GLE) including a non-Markovian term, and the Empirical Model Reduction (EMR) approach to derive GLEs – for various synthetic and real-world time series.

In the first part of this chapter, corresponding numerical simulations of all three inverse methods (KM, MZ, and EMR) were examined in terms of PDFs and ACFs of the simulated time series, as metrics for assessing the models' performance. We generally observed a nearly optimal performance of all three approaches for unimodal Markovian systems. For non-Markovian systems, the MZ and EMR strongly outperformed the KM approach as theoretically expected. According to our results, the performance of

SDEs strongly relies on the effects of memory on the underlying dynamics. We could show that LEs (as derived by the KM approach) obtain better results for systems with weak memory contribution but asymmetric multiplicative noise since they can directly estimate the state-dependent noise with higher precision. On the other hand, for systems with memory effects and colored-noise forcing, it is essential to consider the non-Markovian closure terms. Hence, the MZ and EMR approaches can be regarded as more reliable in reconstructing the dynamics of systems exhibiting strong memory effects. In these two methods, the interactions between observed and unobserved variables are taken into account in terms of memory effects. That is, the EMR approach incorporates adaptive numbers of memory steps, while the MZ method considers the full memory of a system via the kernel K.

In the second part, using a data-driven KM approach, we analyzed the stability landscape of the δ^{18}O and dust recordings in one and two-dimensional frameworks. We found that, in the decoupled setting, the δ^{18}O displays mono-stable dynamics, whereas the dust record displays bistability features. An abrupt transition in the δ^{18}O could then entail a regime switch in the atmospheric configuration. We observed non-vanishing higher-order KM coefficients in the δ^{18}O suggesting that the dynamics of δ^{18}O, unlike the dynamics of the dust record, cannot be modeled as a purely continuous stochastic process. We observed that, even though trajectories of both proxies look similar, the reconstructed potential from drift dynamics are different. To explore the origin of this discrepancy, we analyzed the coupled system of δ^{18}O and dust records. In the two-dimensional setting, we found the position of δ^{18}O's stable fixed is controlled by the current value of the dust, suggesting that couplings between the two are indeed highly relevant.

5. Predicting Climate Variability Using Machine Learning Approach

After reconstructing the underlying dynamics of complex systems in the previous chapter, we now continue to forecast their behavior using ANN approach. The main focus of this chapter is to develop a method based on a recurrent neural network, called Echo State Network (ESN) (see section 2.3.1) that can reliably predict non-deterministic time series with comparably low signal-to-noise ratios, which frequently arises in climate science. In this study, we will apply our developed algorithm to three different climate oscillations, the El Nino/Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), and the Atlantic Multidecadal Oscillation (AMO). These oscillations are associated with large-scale fluctuations in air pressure, sea temperature, and wind direction that can substantially impact global weather patterns. Therefore providing skillful forecasts of these climate modes are essential for society and policymakers. Notably, the predictive skill of the conventional ESN algorithm is significantly affected by the intrinsic characteristic of climate data (i.e., low signal-to-noise ratio). Exploiting the peculiar aspects of climate variability in which the slow mode can be perturbed by high-frequency forcing (e.g., westerly wind bursts or the Madden- Julian Oscillation), we estimate the future behavior of high-frequency variability from its past history using PNF method [30] (introduced in section 2.3.2). By incorporating such information into the ESN model, we aim to enhance its ability to forecast the target time series for a longer lead-time.

The chapter is organized as follows: we first briefly introduce the three climate oscillations used in this study, then we explain the methodological setup for conducting the analysis. Finally, the feasibility of our approach will be evaluated based on the predictive skill of the resulting model. The results presented and the figures shown in this chapter closely follow the publications (Hassanibesheli, F., Boers, N., and Kurths, J. (2021). Long-term ENSO Prediction with Echo-State Network. Environmental Research: Climate, in review) and (Hassanibesheli, F., Boers, N., and Kurths, J. (2022). Predicting Climate Oscillations Using Echo-State Network, in preparation).

5.1. Large-Scale Climate Index Data

5.1.1. El Niño-Southern Oscillation

The El Niño-Southern Oscillation (ENSO) is the dominant variability mode of the global climate system on interannual time scales originating in the Tropical Pacific Ocean [173]. This strong climate variability is generated through the atmosphere and ocean interactions and can impact climate patterns in various parts of the world (e.g., associated with floods and droughts). ENSO can be classified into three main variability modes, namely the warm (i.e., El Niño) and cool (i.e., La Niña) phases with sea-surface temperature (SST) anomalies substantially above and below average, respectively, as well as the remaining neutral phases [161]. The warm ENSO phases have typical yet irregular return periods between 3 and 7 years, rendering them challenging to predict [210]. In the neutral phase, the trade winds blow west resulting in a pile-up of warm water masses at the western boundary of the tropical Pacific Ocean.

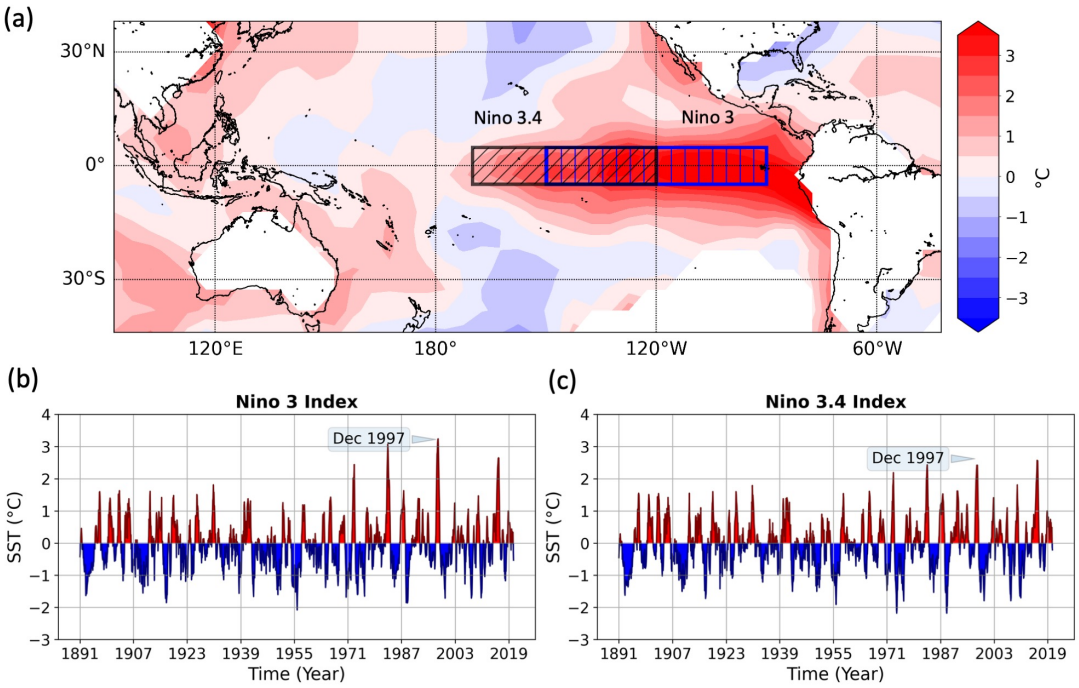

FIGURE 5.1.: Monthly sea-surface temperature (SST) anomalies in the Pacific ocean for the 1997 El Niño event, and SST-based ENSO indices. **(a)** SST anomalies [99] in the Pacific ocean for December 1997, known as the strongest El Niño on record. The blue and black boxes delineate the Niño 3 (5N-5S, 150W-90W) and the Niño 3.4 (5N-5S, 170W-120W) regions from where SST anomalies are used to define the Niño-3 and Niño-3.4 indices [60], respectively. **(b)** The Niño-3 index, given by a time series of SST anomalies averaged over the Niño-3 region in the eastern tropical Pacific as shown in (a), from 1890 to 2019. **(c)** SST anomalies averaged over the Niño-3.4 region from 1890 to 2019. Red and blue colors indicate SST anomalies above and below zero, respectively.

The resulting east-west SST gradient causes air to ascend in the western Pacific and circulate back to the eastern boundary of the tropical Pacific Ocean, where it descends again. This atmospheric circulation system is called the Walker cell. During El Niño (La Niña) phases, this circulation is weakened (strengthened), leading to warm (cool) SST anomalies in the central and eastern parts of the tropical Pacific. The predictability of ENSO at interseasonal and longer time scales has attracted substantial attention, using process-based general circulation models [70, 102] and statistical approaches [91, 108, 170]. Since ENSO dynamics is arguably nonlinear [7], also nonlinear statistical models have been introduced [30, 108]. In addition to these data-driven inverse modelling approaches, several statistical forecasts of El Niño events based on complex network theory have been proposed. Nevertheless, as for the classical statistical approaches above, the forecast horizon of network-based approaches remains limited to one year[132, 149]. Here, we employ monthly SST anomalies for the period 1890-2019 [60], averaged over the Niño-3 (5N-5S, 150W-90W) and Niño-3.4 (5N-5S, 170W-120W) regions in the tropical Pacific, which are commonly used to define the Niño-3 (Fig. (5.1b)) and Niño-3.4 indices (Fig. (5.1c)).

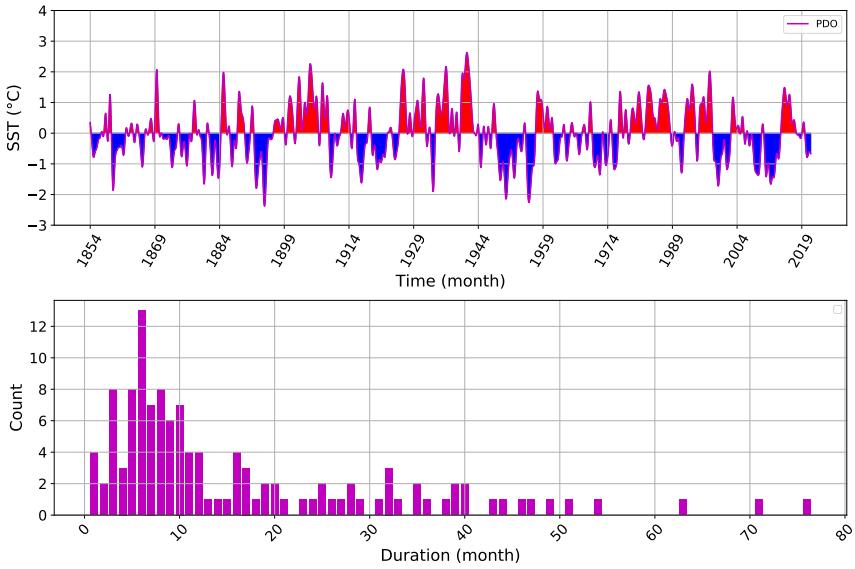

FIGURE 5.2.: Pacific Decadal Oscillation (PDO) index time series during 1854-2019, with warm and cold phases illustrated in red and blue, respectively. The lower panel displays the distribution of PDO phase lengths.

5.1.2. Pacific Decadal Oscillation (PDO)

In the second case study, we focus on the prominent mode of decadal climate variability in the Pacific Ocean known as Pacific Decadal Oscillation (PDO) [140, 157]. PDO has been first identified in the late 1990s [141] as the leading empirical orthogonal function (EOF) of monthly North Pacific SST variability. It represents a long-lived El Niño-like pattern in the North Pacific basin that can cause decadal-scale increases in drought and heavy rainfall frequency over the United States, Canada, and Siberia through atmospheric teleconnections [211]. The PDO comprises two modes, the warm phase, which is associated with negative SST anomalies in the central and western regions of the North Pacific, and the cold phase, during which the opposite pattern occurs. The amplitude of the PDO reaches its climax during November to June and then strongly drops throughout the late summer-early autumn. Multiple processes of different origins may drive PDO variability on different time scales. For instance, ENSO variability is known as the driving factor for the PDO on the interannual time scale through the "atmospheric bridge" [200]. Another candidate mechanism is related to the regional ocean-atmosphere interactions over the mid-latitudes in the North Pacific [218]. Remote forcing from the western Pacific and the eastern Indian Ocean is another relevant driving process [95]. Some other research on the PDO demonstrated the impact of stochastic forcing from internal atmospheric dynamics on North Pacific SST anomalies [221]. Different approaches from stochastic models [188] and linear inverse models [156] to fully coupled climate models [147, 195] have been introduced to identify and predict the underlying dynamics of PDO variability. Although mechanisms contributing to the PDO are relatively well-understood, tracing out the actual evolution of this phenomenon remains challenging. A possible explanation for this complication is that PDO variability emerges as a manifestation of interactions of multiple phenomena on different timescales with different forces. For instance, it has been shown that under the influence of greenhouse-gas-induced warming conditions [122], the PDO prediction skill of coupled general circulation models drastically decreases. It has been demonstrated that greenhouse warming forces can prompt Rossby waves to propagate faster, which consequently results in decreasing the amplitude of the

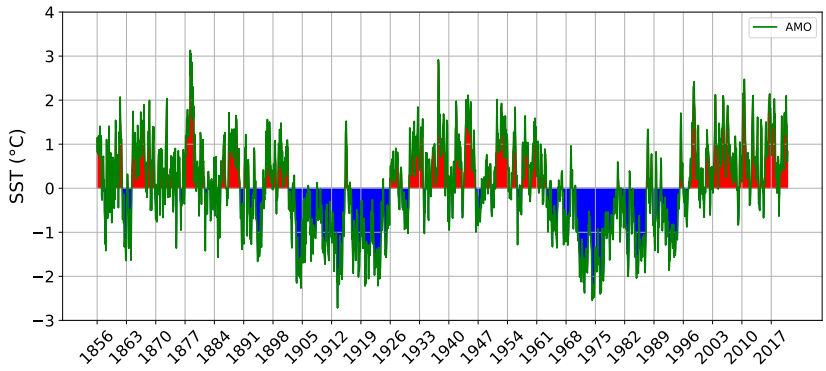

FIGURE 5.3.: Atlantic Multidecadal Oscillation (AMO) index time series. Warm and cold phases are illustrated in red and blue respectively.

PDO [122]. The PDO index used in this study is a monthly SST anomaly poleward of 20N in the Pacific basin from 1854-2019 (see Fig. (5.2)).

5.1.3. Atlantic Multidecadal Oscillation (AMO)

As the final case study, we apply our model to predict Atlantic Multidecadal Oscillation (AMO) index. AMO is a significant component of climate variability in the North Atlantic basin constructed based on average sea surface temperature anomalies (SST). By changing the air temperatures and rainfall, the AMO can trigger prominent decadal-scale climate anomalies, including drought in Africa, severe hurricanes in northern America regions [85], and variability in northeast Brazilian rainfall [55]. It has also been shown that AMO can significantly affect the Indian monsoon rainfall [121] that consequently impacts the southern part of Central Asia. The transition between warm and cold phases of AMO occurs every 20–40 years [4]. During AMO's warm phase, more hurricanes occur in the Atlantic region. While its cooler phase is associated with reduced rainfall in the Sahel region of Africa. Due to AMO's multidecadal time scale and intricate interactions of various climate sub-components, understanding possible forcing mechanisms driving the AMO are under considerable debate. According to one school of thought, based on climate model simulations, Atlantic Meridional Overturning Circulation (AMOC) [39] is the primary driver of AMO, which is associated with internal ocean variability. On the other hand, others argue that external forcing such as volcanic eruption and solar variability contribute to the multidecadal SST variability and are key drivers of AMO [105, 162]. Some other processes, such as changes in the strength of the ocean gyres and Gulf Stream or decadal fluctuations in sea ice, can also contribute to decadal variability over the Atlantic [40]. During the last decades, substantial progress has been made to model and predict the temporal structure of the AMO using different methods from coupled global atmosphere–ocean models [33] to probabilistic approaches [201] and statistical methodologies [220]. In this study we use the unsmoothed monthly AMO index calculated from the Kaplan SST dataset [50] that covers the period from 1856 to 2020.

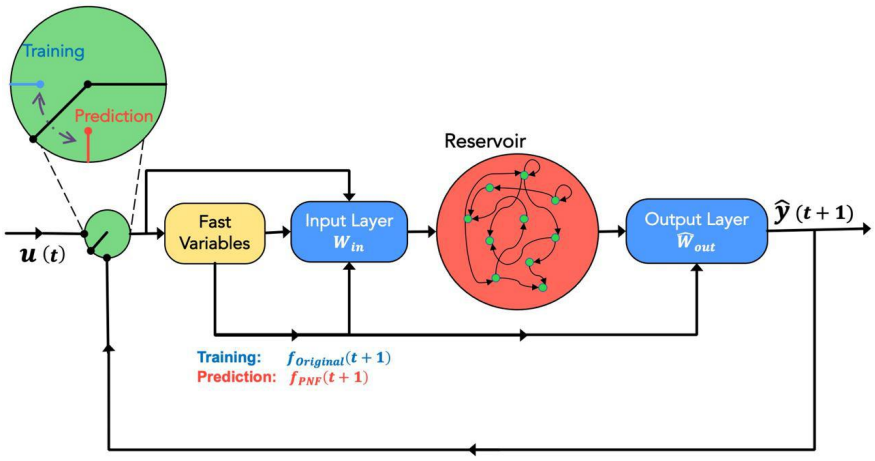

FIGURE 5.4.: Schematic diagram of ESN-PNF setup.

5.2. Method

5.2.1. ESN Implementation and Combination with PNF method

Our proposed ESN implementation consists of two main steps; the training phase during which a loss function is minimized to find optimal output weights, and the prediction phase during which the optimized ESN is used to predict unseen data.

Our approach's fundamental assumption is that climate variability can be disentangled into high- and low-frequency modes, where the dynamics of low-frequency variability is under the influence of high-frequency forcing. Several methodologies and techniques have been developed to decompose time series into slow and fast variability components, such as, e.g., singular spectrum analysis (SSA) [74, 209], spectral methods such as the Butterworth (BW) filter [62, 182], or moving average (MA) filters [6]. Upon a careful evaluation of their performance, we employ the BW filter to decompose the training data into a low- and high-frequency component. The latter is treated as noise forcing. We then pass both components as input to the ESN. At each time, we perturb the low-frequency component at time t by noise at time $t+1$. We train the ESN based on 80% of the original dataset which is prior to t^*, the time that we aim to commence the prediction task (see Fig. (5.4)). The weights of the output matrix W_{out} (see Eq. 2.51) are then optimized by minimizing the loss function given by the root-mean-squared error (RMSE) between the values of the actual time series confined to the training interval and the corresponding simulated values, using ridge regression [80]. For an appropriate ESN, numerous iterations with random initializations must be carried out to tune the hyperparameters (see section 2.3.1.1).

Then, using the optimized W_{out}, the trained model starts recursively forecasting L consecutive time steps ahead of t^*. In the prediction phase, since we do not have any information about the future of the noise, we apply the Past-Noise Forecasting (PNF) method (for further details, see section 2.3.2). Based on perturbation techniques, this approach attempts to forecast the actual path of the future noise from the past noise segments. Within the framework of PNF, we select noise segments with length $L = 25$ where the phase of the system resembles the one just preceding the state at t^*. To find an analogous phase, we look at the difference between $t^* - \Delta$ and $[t_j, t_j + \Delta]$ in the smoothed version of the target time series obtained from BW low-pass filter. We consider these two segments analogous if $\alpha \leq 0.5$ and $\gamma \geq 0.9$ (see Eq. 2.53). Here, respectively, α and γ denote RMSE and PCC threshold values. Depending on the

system of interest, these two parameters must be tuned accordingly to avoid an empty set. Once the selection criteria are fulfilled, we can determine noise segments from the time series of the noise $[t_j, t_j + L]$ that potentially can drive the system into the future from t^* to $t^* + L$. The prediction for the first time step is then, together with corresponding candidate noise analogues, passed to the ESN as inputs to predict the second time step, and so on. We repeat these training and prediction procedures for each time step i within the prediction interval. Finally, we can extract the predicted time series based on different lead times ranging from 1 to 25 time steps.

5.3. Results and Discussion

5.3.1. ESN-PNF Performance on Predicting ENSO Indices

In the following, We focus on forecasting the Niño-3 and Niño-3.4 indices (Fig. (5.5)) using the ESN-PNF model during the period of 1982-2019. To evaluate the skill of the proposed model, we consider four different metrics, namely (I) the Root Mean Square Error (RMSE) and (II) the Pearson correlation coefficient (PCC) to evaluate the overall performance in predicting the ENSO index, as well as (III) the Heidke-Skill Score (HSS) and (IV) the probability of detection (POD) to evaluate the binary prediction of El Niño events (for more details see Appendix C.1.1). For Niño-3 (Niño-3.4) the RMSE remains almost constant at around 0.5°C for lead times between one and 14 months and then increases linearly up to about 1.4°C at 24 months lead-time (Fig. (5.5a)). The PCC between the observed and simulated time series for the prediction phase remains constant around 0.8 also up to lead times of around 14 months and drops below 0.5 after 18 months (Fig. (5.5b)). To better evaluate the performance of our ESN-based model in comparison to the existing predictive models, we depict the ENSO correlation skill of the previously introduced CNN [72] and some dynamical forecast systems [36, 136, 150]. It can be observed from (Fig. (5.5b)) that our model is superior to dynamical systems at lead times longer than six months. For instance, the correlation skill of the Niño-3 index in the ESN model (red) is above 0.5 for a lead time of up to 18 months, while it is 0.35 at a lead of 18 months in the SINTEX-F5 [136](green). Note that, following [72] we consider forecasts as skillful if the correlation coefficients are above 0.5. To assess the skill of our model in predicting El Niño events (i.e., time steps with ENSO index larger than one), we applied the Heidke skill score (HSS) and the the Probability of Detection (POD) (for more details see C.1). According to panels c and d in Fig. (5.5, both event-based metrics show high skill up to forecast lead times of 18 months. Our results remain similar when 5-month running averages of the Nino-3 index are considered and when El Niño events are defined as periods for which the smoothed index is above 0.5° for six consecutive months.

Generally, the performance of a predictive model depends on the length of the provided training data. One of the fundamental questions during training NNs is how much data is required to reasonably approximate the unknown mapping function from input to output. It is common knowledge that training a model on larger volumes of data can result in better performance since it can capture the inherent features and dependencies more efficiently. In Fig. (5.6), we demonstrate how the size of the training dataset can directly affect our model's prediction skills. It can be observed that with increasing the size of the training data, the forecast skill of the ESN-PNF model increases (i.e., higher values in terms of PCC, HSS, and POD). Additionally, similar to Fig. (5.5), we evaluate the overall performance of the ESN-PNF model trained on longer data from 1981-1991 (see Fig. (5.7)). We can see that the model achieves an even higher skill during the first 14 months compared to the training data from the interval 1891-1982.

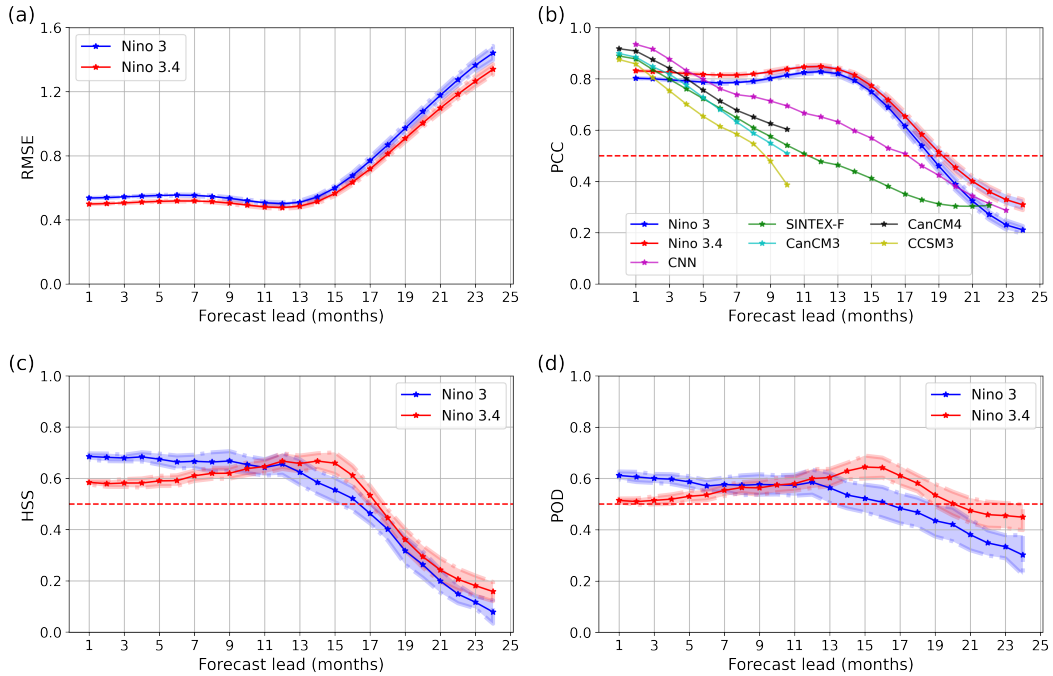

FIGURE 5.5.: Summary of the ESN-PNF forecast skill for the Niño-3 and Niño-3.4 indices. Panels **(a)** and **(b)** show the prediction skill of the model in terms of the Root Mean Square Error (RMSE) and the Pearson correlation coefficient (PCC) between predicted and observed values, respectively. The model is trained with a fixed training data length $T = 1092$ and the prediction task starts from 1982. Results for Niño-3 are shown in blue and results for Niño-3.4 in red color. In panel **(b)** the magenta curve shows the ENSO correlation skill of a previously introduced CNN[72] together with a comparison to other process-based based ENSO predictions. The ability of our ESN-PNF model to predict El Niño events (i.e., times with Niño-3 index above 1 $°C$) are assessed by two binary classifiers, the Heidke Skill Score (HSS) and the Probability Of Detection (POD), in **(c)** and **(d)**, respectively. Solid lines indicate the average over 100 realizations of the forecast, and shading around the lines represent $\pm 1\sigma$.

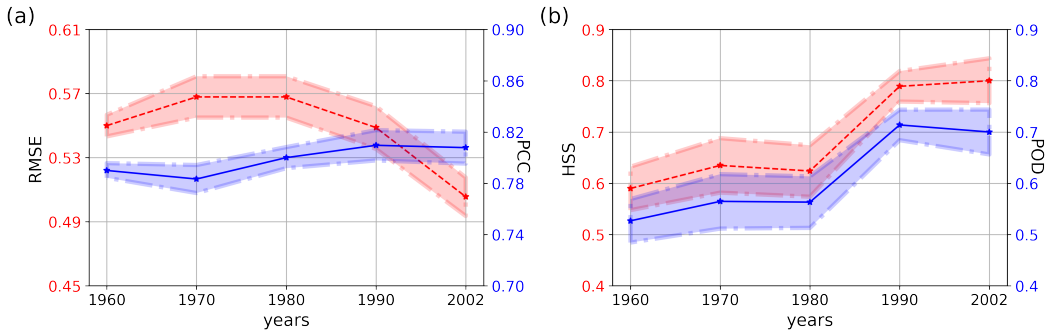

FIGURE 5.6.: Dependency of the forecast skill on the training length. Panel **(a)** illustrates the RMSE and Pearson correlation coefficient (PCC) values between the predicted and absolute values of Niño-3 index at a lead of 12 months with respect to different starting points. Panel **(b)** shows HSS and POD for the corresponding lead time. Our ESN model exhibits overall better performance the longer the training data. Shading around lines represents $\pm 1\sigma$.

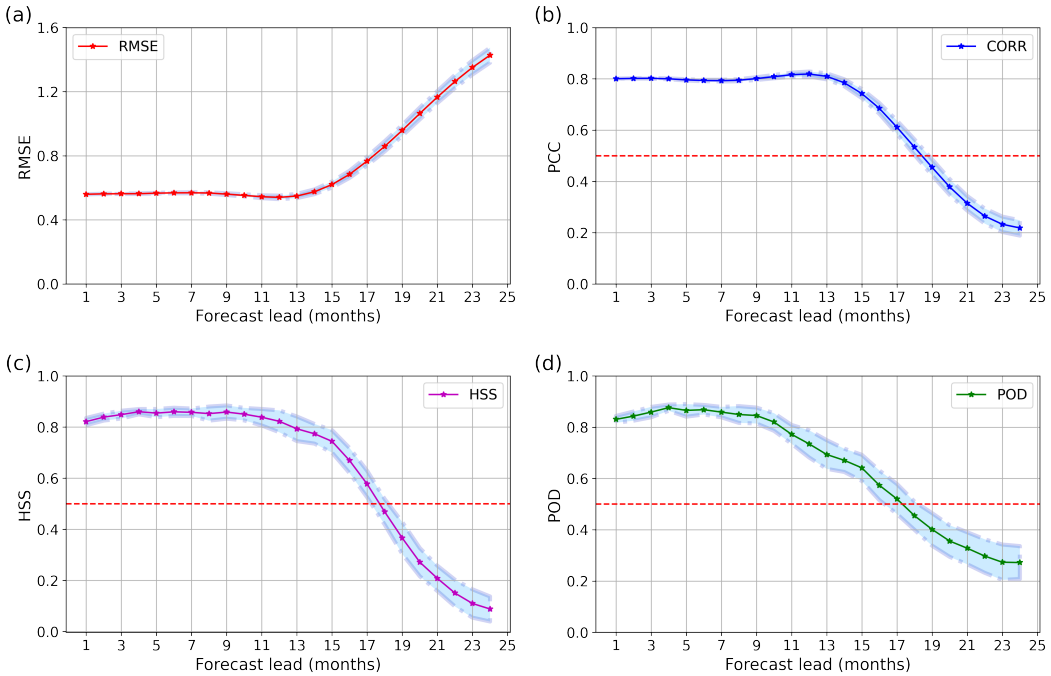

FIGURE 5.7.: Summary of the ESN-PNF forecast skill for the Niño-3 index. Panels **(a)** and **(b)** respectively show the prediction skill of the model in terms of Root Mean Square Error (RMSE) and Pearson correlation coefficient (PCC) between predicted and observed values. The ability of the model to detect El Niño events (i.e., times with Niño-3 index above 1 $°C$) are assessed by two binary classifiers, the Heidke skill score (HSS) and the Probability of detection (POD), in **(c)** and **(d)**, respectively. Solid lines indicate the average over 100 realizations of the forecast, and shading around lines represent $\pm 1\sigma$. The results correspond to the reservoir setup trained with a fixed training data length $T = 1212$ and the prediction starts from 1992.

To further evaluate the reliability of our ENSO forecast, we investigated the sensitivity of the model's forecast skill to the target months at different lags, see Fig. (5.8). In agreement with previous study [72], our ESN model exhibits the longest forecast horizon for target months in boreal winter (the correlation between the observed and simulated ENSO index is above 0.5 for 21 months) and the shortest forecast horizon (18 months) for target months in late boreal spring. This spring predictability barrier [215] may be a result of the comparably weak Walker circulation and susceptibility of the coupled ocean-atmosphere system to the external forcing at that time of the year. Nevertheless, a recent CNN model [72] has a valid forecast horizon (i.e., lead times at which the correlation remains above 0.5) substantially longer than process-based dynamical models (11 months vs. 4 months) for the May-June-July season. For the May–June–July season, the valid forecast horizon of our ESN-PNF model is even substantially longer compared to the CNN model(18 months vs. 11 months). This indicates that our method is less affected by the spring predictability barrier than the CNN and process-based models, possibly because we estimate the future fast variables that can drive the system. As an example, the predicted SST anomaly between 1983 and 2019, at the 18-month lead-time, shows a high visual resemblance to the observational Niño-3 index (Fig. (5.9a)), although this is the longest lead time for which we consider our forecast to exhibit skill. The predicted values are computed as averages over 100 realizations at the 18-month lead-time. We also examined the

performance of our ESN-PNF model by comparing the statistical properties such as the auto-correlation functions (ACFs, Fig. (5.9b)) and probability density functions (PDFs, Fig. (5.9c)) of the predicted and observational time series. The ACF shows a very close resemblance between the observed and simulated indices in terms of their correlation structure, and according to the Kolmogorov–Smirnov test, the hypothesis that the underlying distributions of the observed and simulated indices are identical cannot be rejected ($p = 0.35$).

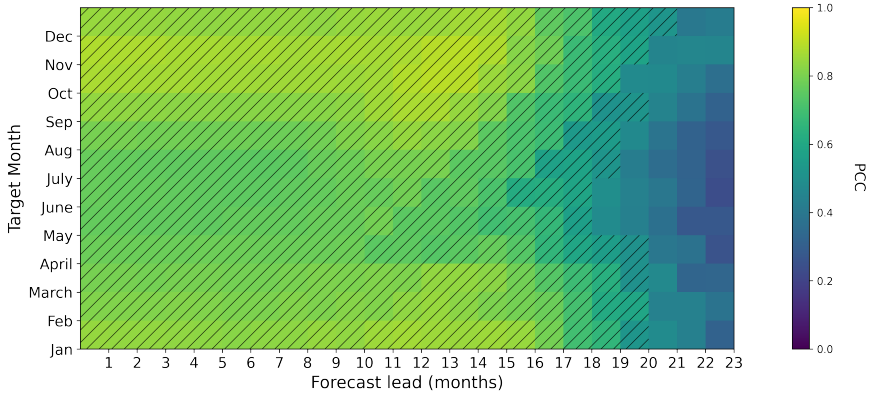

FIGURE 5.8.: Correlation skill of the ESN-PNF model as function of lead time and target month for the Niño-3 index. The Pearson correlation coefficient (PCC) between the predicted and observed Niño-3 index targeted to each calendar month at different lead times, averaged over 100 realizations. Here we train the model on a fixed training data length $T = 1092$ and start forecasting from 1982.

Note that the prediction accuracy of our ESN-PNF model varies moderately over time, depending on the absolute values of the employed ENSO index. In Fig. (5.10) we have analyzed the twelve-month running mean of the Niño-3 index from 1983 to 2018. We can observe higher RMSEs during the 1997-98 and 2015-2016 El Niño events when the maximum SST anomalies reaches about 3.5° in both cases. Hence, for El Niño and La Niña events with particularly high or low values of the ENSO index, the RMSE is comparably higher. Additionally, the performance of our ESN-based prediction scheme can be slightly affected by varying the cutoff value C of the low-pass filter used to separate the low- from the high-frequency components of the ENSO index. We found that $C = 0.03$ is the optimal choice of the cutoff value in terms of the overall forecast skill (see Fig. (C.1)).

Further, we evaluated the ESN-PNF performance in predicting the Niño-3.4 index and compared its predictive skill with the SINTEX-F dynamical model [72] in terms of PCC. Fig. (5.11a) displays the PCC between the predicted and observed Niño-3.4 index targeted for each calendar month at different lead times. We trained the model on a fixed training data length $T = 1092$ and started forecasting from 1984 to 2017. It can be observed from Fig. (5.11b) that the forecast skill targeting the May–June–July season has a correlation skill above 0.5 only up to a lead of four months in the SINTEX-F dynamical model, while the ESN-based model shows a higher correlation skill for almost all targeted seasons. Additionally, we compared the predicted time series of the December–January–February (DJF) Niño-3.4 index for the 17-month-lead obtained from our model with the CNN and SINTEX-F dynamical models. It can be observed from Fig. (5.12) that the ESN-based model (blue) correctly predicts the ENSO amplitude and exhibits better performance. We further investigated the statistical properties of the predicted and observed Niño-3.4 index for a 17-month-lead forecast using the ESN-

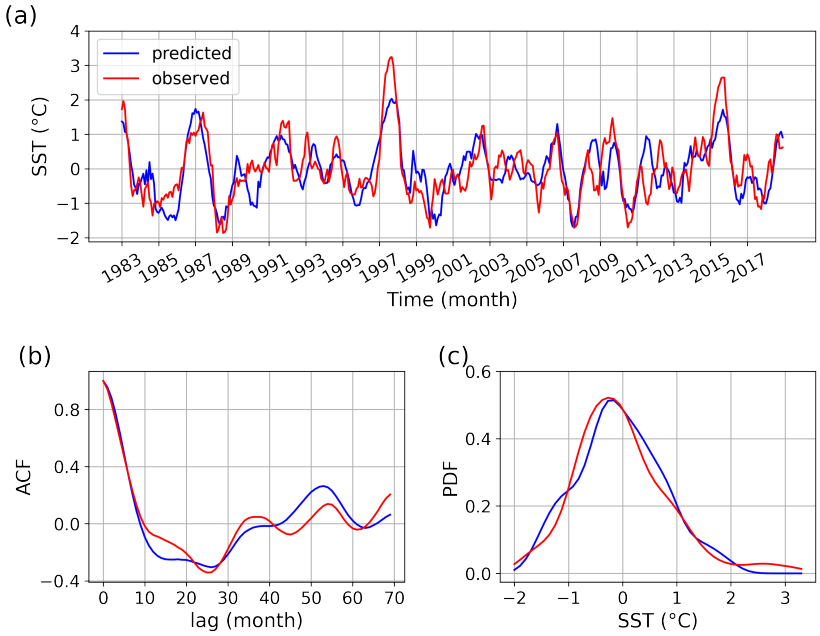

FIGURE 5.9.: Comparison between the statistical properties of the predicted and observed Niño-3 indices. **(a)** Time series of SST anomalies of the Niño-3 index (red) and the predicted time series (blue) at the 18-month lead-time. The model is trained with a fixed training data length $T = 1092$ and the prediction task starts from 1982. **(b)** and **(c)** compare the statistics (ACF and PDF) of the original and predicted time series; for the predictions, averages over 100 realizations at the 18-month lead-time are taken.

based model (see Fig. (5.13)). Similar to the Niño-3 index, our model can successfully follow the underlying statistical properties of the Niño-3.4 index.

5.3.2. ESN-PNF Performance on Predicting PDO Index

In contrast to ENSO, there have been relatively fewer efforts to forecast PDO variability using ANNs [64]. Here, we proceed with an analogous analysis conducted in section 5.3.1 by employing the ESN-PNF model on the monthly PDO index. Some of the previous studies [119, 131] applied low-pass filters with 5 to 11-year cutoffs to filter out higher interannual frequencies. However, we try to predict the temporal evolution of the monthly PDO index using the information of both low and high-frequency variability. To learn the underlying features of the PDO index, we trained the ESN-PNF with a fixed training data length of $T = 1450$ months and started prediction from 1981 onward until 2019.

As we mentioned in section 5.3.1, the performance of our model can be influenced by varying the cutoff value of the low-pass filter. To investigate that, we measured the RMSE and PCC between observational and predicted data at different cutoffs. We observed that the most extended forecast horizon can be obtained at $C = 0.03$ (months^{-1}) (see Fig. (C.3) and Fig. (C.4)). As it can be observed from Fig. (5.14a), the RMSE at this cutoff remains almost constant at around 0.5°C for lead times between 1 to 10 months and then increases linearly up to about 1.8°C at 24-month lead-time. The corresponding PCC remains constant at around 0.8 also up to lead times of around 10 months and eventually drops below 0.5 after 18 months (see Fig. (5.14b)). Compared

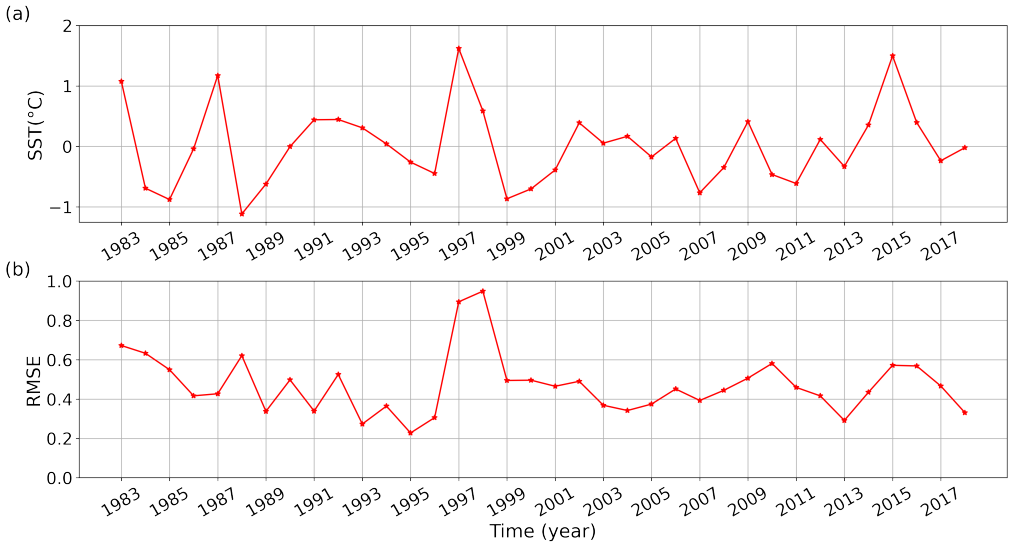

FIGURE 5.10.: Dependence of the forecast skill on the target year for the Niño-3 index. **(a)** Twelve-month running mean of the Niño-3 index from 1983 to 2018, here the training length is equal to 1092 . **(b)** Forecast skill of our ESN model from 1983 to 2018 in terms of the RMSE. Each yearly value is the average RMSE over all months of the corresponding year, calculated on 12-month lead. We note that in some years with exceptionally strong El Niño events, the prediction accuracy of the model is affected slightly.

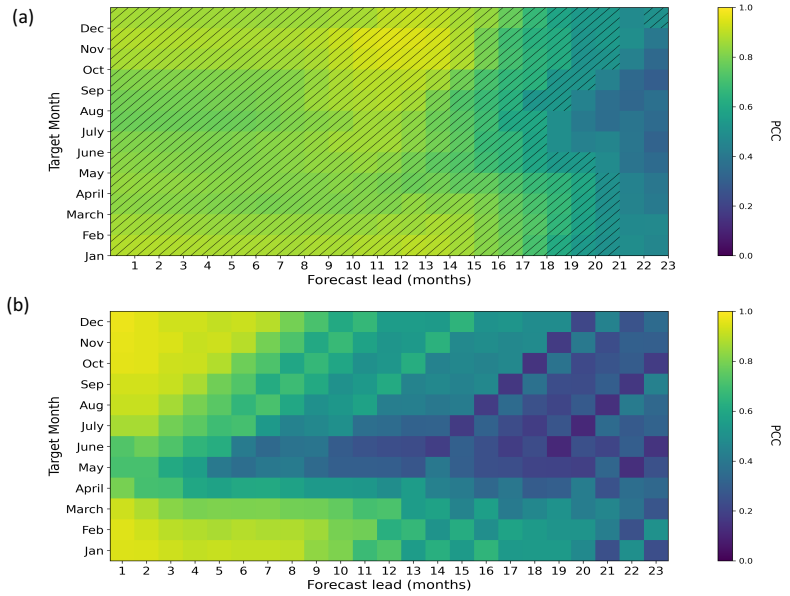

FIGURE 5.11.: Same as (Fig. 5.8) but for Niño-3.4 index. **(a)** The Pearson correlation coefficient (PCC) between the predicted (obtained from ESN-PNF) and observed Niño-3.4 index targeted to each calendar month from January to December, at different lead times. Hatches indicate combinations of target months and lead times for which the correlation of observed and predicted ENSO index is above 0.5. **(b)** The correlation skill of the Niño-3.4 index in the SINTEX-F dynamical forecast system [72]. The prediction period for both models is between 1984 and 2017.

to the model proposed by Gordon et.al. [64], our model exhibits better performance.

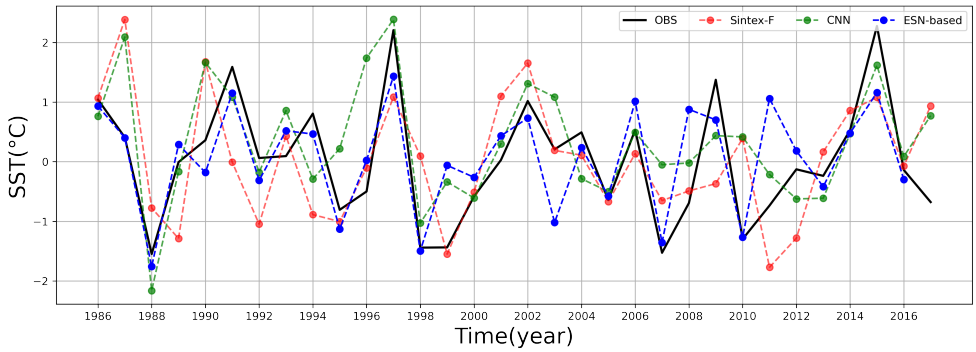

FIGURE 5.12.: Time series of December–January–February (DJF) season Niño-3.4 index for an 17-month lead forecast using the ESN-based model (blue), CNN (green), and the SINTEX-F model (red).

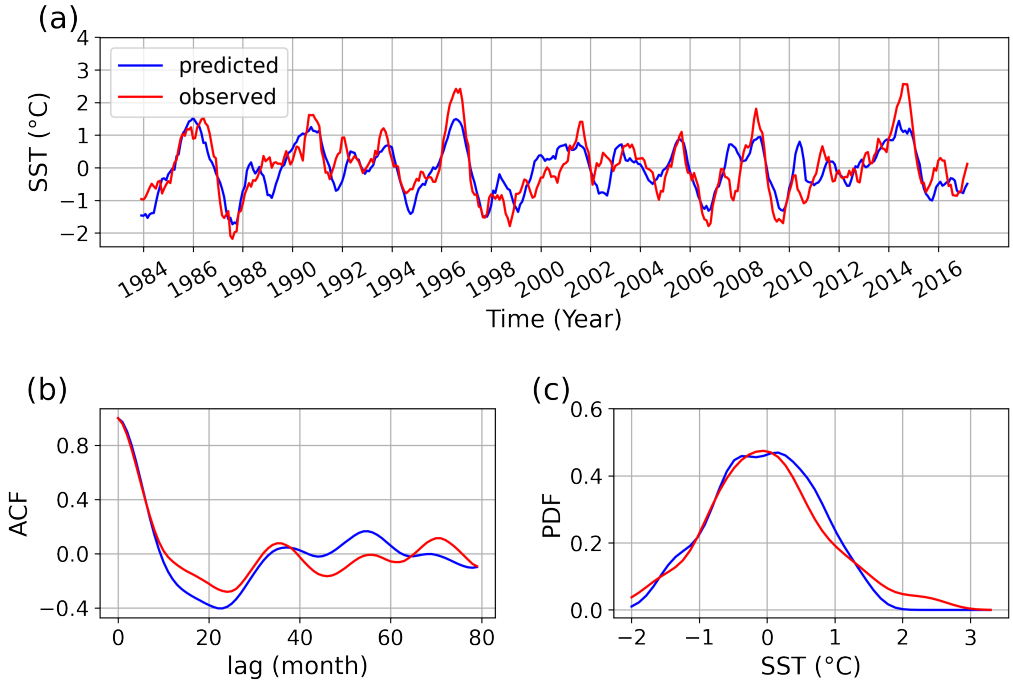

FIGURE 5.13.: Same as (Fig.(5.9)) of the main text, but for the Niño 3.4 index. **(a)** Time series of SST anomalies of the Niño 3 index (red) and the predicted time series (blue) at the 17-month lead-time during 1984-2017. **(b)** and **(c)** compare the statistics (ACF and PDF) of the original time series and predicted time series; for for the predictions averages over 100 realizations at the 17-month lead-time are taken.

Their model, which is based on a single-layer ANN, could only predict PDO for up to 12 months.

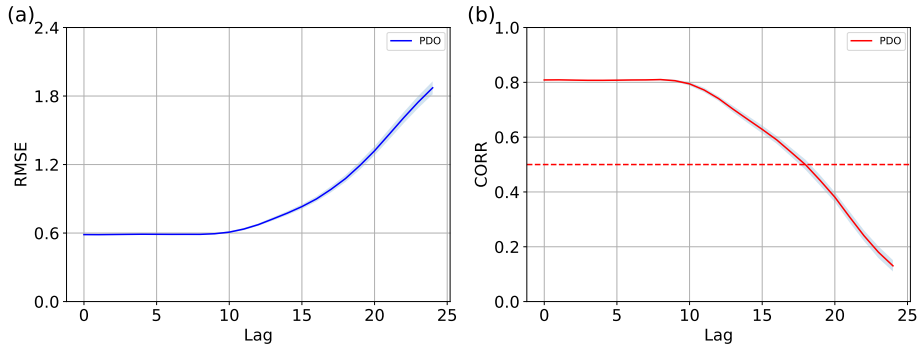

FIGURE 5.14.: Summary of the ESN forecast skill for PDO. Panels **(a)** and **(b)** show the prediction skill of the model in terms of Root Mean Square Error (RMSE) and Pearson correlation coefficient (PCC) between predicted and observed values.

Finally, in Fig. (5.15), we compared the statistical properties of the predicted and observational data at the 18-month lead-time in terms of the ACF and PDF. In particular, Fig. (5.15a) shows that most of the transitions between positive and negative phases can be captured by the model, even at the 18-month lead-time (see also Fig. (C.6)). Even though the PCC stays above 0.5 up to 18 months, the statistical properties of the predicted time series can not properly follow the original PDO index as we expect. That may be due to the low signal-to-noise ratio of the PDO index, which can decrease the ability of the ESN-PNF to predict accurately at a longer lead-time. Comparing the PDO and ENSO indices in terms of spectrum frequencies (see Fig. (C.2)), we can observe that the PDO has a high amplitude at higher frequencies.

5.3.3. ESN-PNF Performance on Predicting AMO Index

In the final case study, we use ESN-PNF to predict the monthly AMO index during 1969-2017. Similar to the previous cases, we evaluated the performance of our model with different smoothing levels in terms of RMSE and PCC (see Fig. (C.5)). We observed that our model achieves better performance at low-pass cutoff $C = 0.03$ (months^{-1}). According to Fig. (5.16b), the correlation skill of the AMO index in the ESN-PNF model exceeds 0.5 for lead times up to 16 months. Additionally, we compared 6-,9-,12-, and 15-month-lead AMO index prediction to evaluate the ESN-PNF performance at different lead times effectively. Conspicuously, the higher the lead-time, the poorer ESN-PNF can follow the trajectory of the original index (see Fig. (5.17)). This observation was corroborated by comparing the PDFs of the predicted and corresponding original data. Fig. (5.18) displays the PDFs associated with 6-,9-,12-, and 15-month lead times of AMO index predictions. According to the p-values of the Kolmogorov-Smirnov (KS) test at different lead times, we can see that the predictive skill of the ESN-PNF at 6-, 9-, and 12-months are significantly higher than 15-month lead-time. The p-value at 15-month lead-time is equal to 0.04, indicating that the underlying distributions of the observed and predicted indices are not identical. This poor performance may stem from the complexity of the high-frequency components in the AMO index.

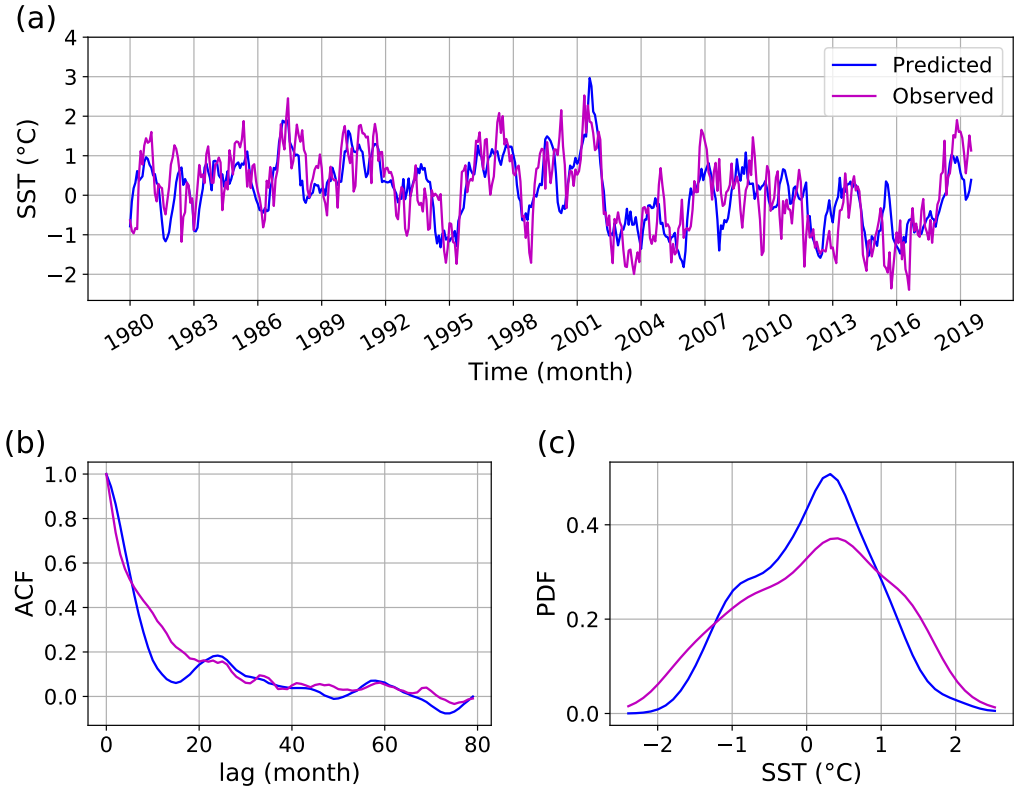

FIGURE 5.15.: **(a)** Time series of SST anomalies of the PDO index (red) and the predicted time series (blue) at the 18-month lead-time. The model is trained with a fixed training data length $T = 1450$ months and the prediction task starts from 1981. **(b)** compare the ACF of the original and predicted time series; for the predictions, averages over 100 realizations at the 18-month lead-time are taken.

5.4. Summary

There have been considerable advances in machine-learning-based approaches to predict SST variability in different ocean basins. One of the critical obstacles most deep learning models encounter in predicting climate phenomena is the unavailability of sufficiently long observational time series. Therefore, deep-learning models must be additionally trained on process-based dynamical model simulations to tackle this limitation. This approach forces deep-learning models to learn the biases and structural errors present in these climate models. In contrast to deep ANNs, the comparably simple ESN we employed here can learn the dynamics of the underlying system from limited amounts of training data. Here, we expanded the ESN approach to predict climate time series. To do so, we combined the ESN model for the low-frequency component with the previously introduced past-noise forecasting (PNF) method, which models the high-frequency component of the time series in question.

In the first case study, we decomposed ENSO indices (i.e., Niño-3, Niño-3.4) into dominant low- and high-frequency variability using low-pass filter techniques and trained the ESN model on the slow mode of the system. To model the effect of the high-frequency forcing on the low-frequency variability, we estimated the potential future high-frequency forcing of the system by relying on the PNF method. This improved predictability shows that interactions across multiple time scales play a crucial role in

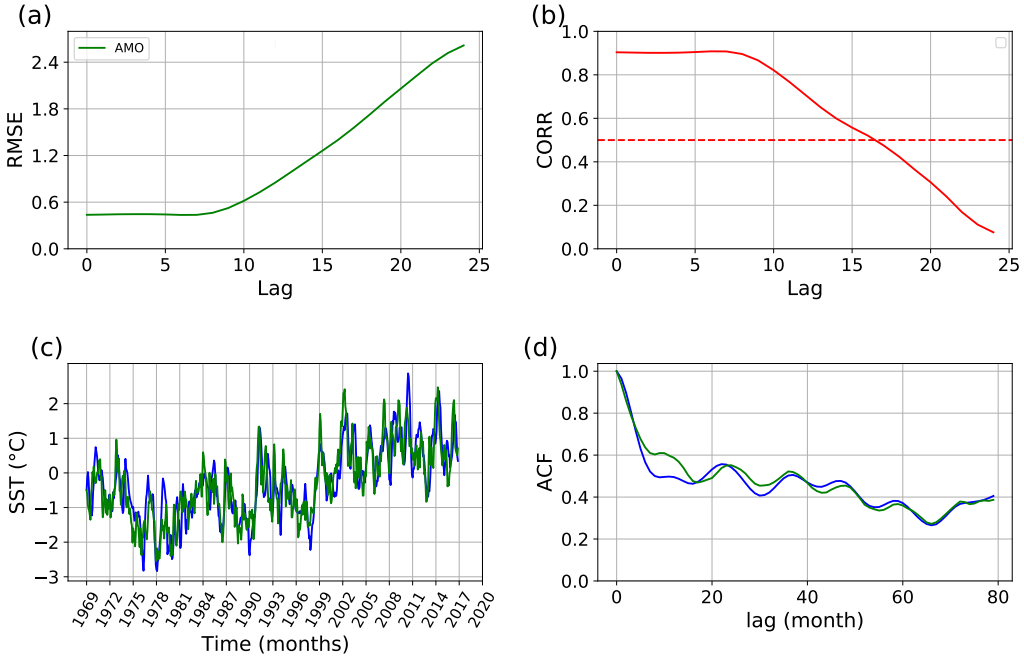

FIGURE 5.16.: Summary of the ESN-PNF forecast skill for the AMO index. Panels **(a)** and **(b)** show the prediction skill of the model in terms of Root Mean Square Error (RMSE) and Pearson correlation coefficient (PCC) between predicted and observed values. Panel **(c)** displays time series of SST anomalies of AMO (green) and the predicted time series (blue) at the 12-month lead-time during 1969-2017. Panel **(d)** illustrates the comparison between the ACF of the original time series and predicted time series.

generating ENSO dynamics. Our ESN-PNF model exhibits high skill in forecasting ENSO variability and predicting El Niño events up to lead times of 18 months, despite the fact that we have trained our model only on single scalar ENSO indices. Our approach outperforms existing statistical, process-based, and deep-learning methods at lead times beyond one year.

In the second and third case studies, we focused on predicting the PDO and AMO indices using the ESN-PNF. Similarly, we showed that our model can capture the critical features of the underlying dynamics of the PDO and AMO. For instance, in the case of monthly PDO, we showed that the predictive skill of our model is much higher than the model presented in [64]. Following the same rationale used in predicting ENSO indices, we demonstrated that cross-scale interactions are also highly relevant for predicting PDO and AMO indices.

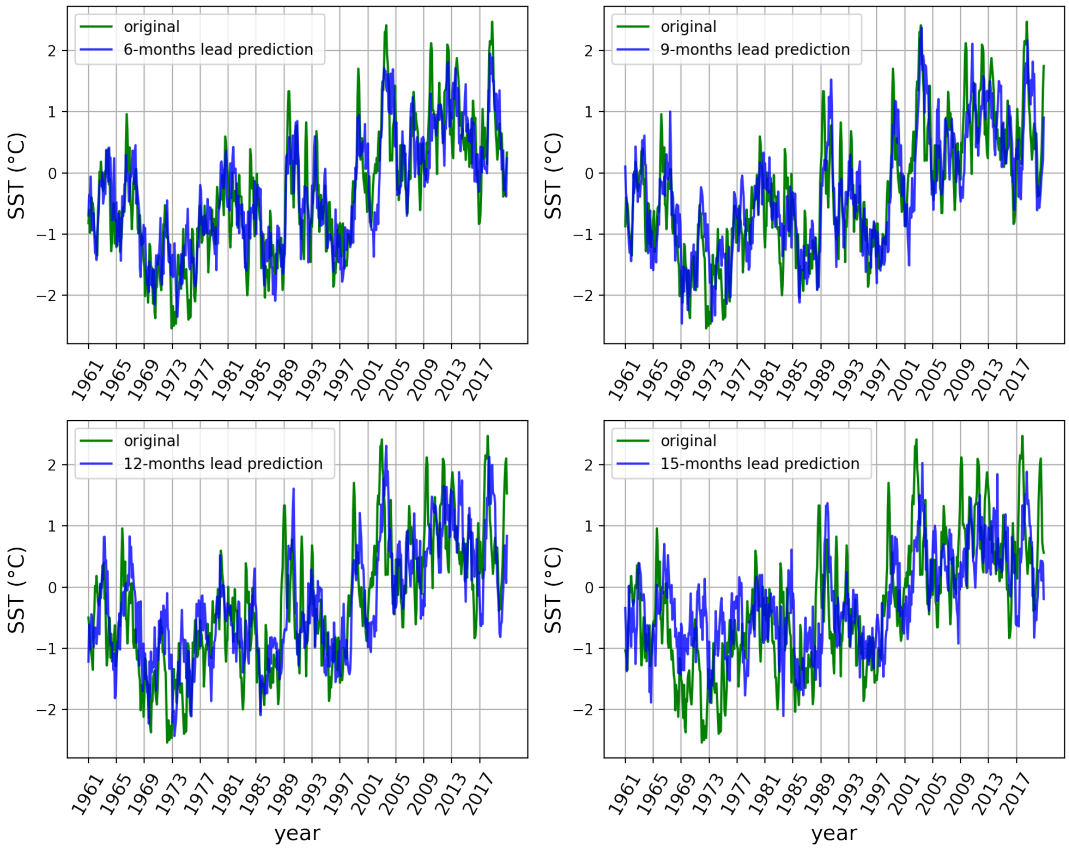

FIGURE 5.17.: Predicted and actual values of AMO index for different lead-times.

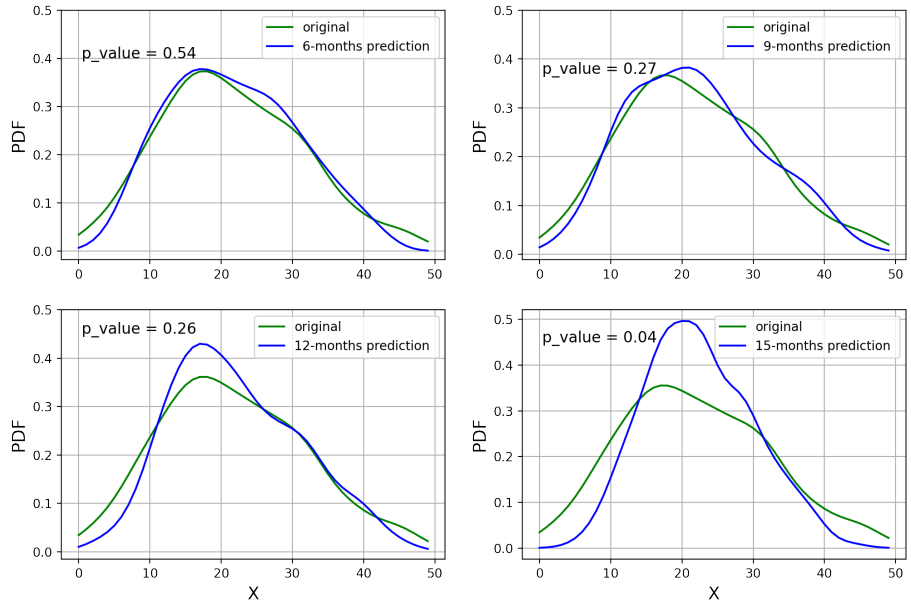

FIGURE 5.18.: PDFs of the predicted and actual monthly AMO index for different lead-times.

6. Conclusions and Outlook

The main objective of this thesis is to advance our understanding of complex systems. Two overarching questions that were investigated are: (I) how one can infer the underlying dynamics of complex systems from given time series and (II) how one can enhance the prediction of future behaviors of such systems. In the following, we review the main findings of the five studies conducted for achieving the mentioned aims and finally finish the thesis with an outlook on future research.

6.1. Conclusion from study 1: Reconstructing Complex Network's Structure from the Timing of Events

In Chapter 3, we have comprehensively compared different strategies to reconstruct the structural connectivity of unknown networks based on their associated functional networks. The functional representations of networks were obtained from statistical associations between the dynamics at each pair of nodes from given event series. To obtain the statistical relationships, we employed event coincidence analysis (ECA) and event synchronization (ES). These two nonlinear similarity measures are commonly used for identifying pairwise statistical associations among discrete event series in neuroscience [174] and climatology [44]. We specifically considered complex networks that exhibit event propagation occurring simultaneously or with some delay from one node to another. To simulate event propagation processes, we studied two different spreading models (i.e., a generic event propagation model and SIRS model) on three different network architectures (i.e., Erdös-Rényi, Barabási-Albert, and Watts-Strogatz networks). In all studied cases, we showed that both similarity measures can successfully capture a vast part of the existing links. However, we showed that in sparse networks, ECA results in a better network reconstruction than ES. On the other hand, in the case of denser networks, we observed that ES compared to ECA performs better in identifying existing connections. Notably, the disparity between the performance of Es and ECA might be related to the characteristics of their coincidence intervals. The coincidence interval in ES is dynamic and local, while it is global and fixed in ECA. Our findings imply that implementing the data-adaptive "dynamic" coincidence interval in ES allows for studying associations at different time scales non-parametrically. Additionally, using the ES for studying synchrony between two event series, which are clustered in time, results in underestimating the number of existing events. This underestimation is due to ES's dynamical coincidence interval shrinkage. Finally, our results revealed that ECA is more suitable than Es when one has prior knowledge about the possible propagation delays. This stems from the fixed coincidence interval of ECA that permits the time scale of interest to be manually determined.

6.1.1. Outlook

In Chapter 3, we employed ES and ECA to capture the statistical similarities between time series of events. However, various alternative approaches can be used to quantify the similarities of regularly sampled event sequences. For instance, one possible strategy would be to use distance measures (e.g., Euclidean or Hamming distances) [159],

which can be easily computed for binary sequences. A second strategy is to interpret individual binary values as symbolic sequences and utilize mutual information as a corresponding similarity measure [198]. Therefore, a systematic comparison of the performance of network inference based on ES and ECA with those measures would be an interesting topic for further studies. Moreover, we considered an unconditional approach, which may lead to overestimating the number of existing links due to indirect linkages (mediated via third nodes). For further investigations, conditional association measures can be employed to disentangle direct from such indirect connections. To the best of our knowledge, only few studies tried to expand the causal network inference developed for time-continuous dynamics [184, 185] to the case of discrete event data [15]. Within the framework of ECA, multivariate and conditional versions have been recently introduced [193] but not yet systematically applied. Future studies should, therefore, focus on the future development of conditional versions of event discrete statistical association measures and their utilization in combination with causal inference algorithms.

6.2. Conclusion from study 2 & 3 : Reconstructing Underlying Dynamics of Complex Systems using Data-Driven Approaches

In chapter 4, we reconstructed equations of motion and corresponding statistical properties of underlying dynamics of complex systems from time series. For this purpose, we employed different reduced-dimension models which characterize the underlying dynamics by isolating a small number of relatively slow degrees of freedom. Using this approach, we could solve the governing equation of motion for different systems using three different stochastic differential equations (SDEs): Langevin equation (LE) [82], generalized Langevin equation (GLE) [38], and Empirical Model Reduction (EMR) [110]. We retrieved the LE and GLE from the Kramers-Moyal analysis (KM) [82] and the Mori-Zwanzig formalism (MZ) [154], respectively. To derive the EMR, we estimated the deterministic term from KM and the stochastic term from data using a recursive procedure. The results in this chapter were divided into two parts.

In the first part, we investigated the performance of three SDEs (i.e., LE, GLE, and EMR) in reconstructing the underlying dynamics of various synthetic and real-world time series. A primary motivation was to comprehensively evaluate the ability of these approaches to solve particular classes of problems. Our results indicated that intrinsic dynamics of systems of interest greatly influence the performance of the resulting SDEs. For instance, statistical properties of systems exhibiting weak history-dependence but strong state-dependence to the noise forcing (i.e., Markov processes) can be approximated better by LE than by the GLE and EMR. In such situations, the LE is of a considerable advantage since it can directly approximate the state-dependent noises. However, we showed that limitations of LE approximation arise in cases where non-Markovian effects are crucial in the system's dynamics. Through extensive analyses of various systems, our results signify that SDEs considering memory effects, i.e., GLE and EMR, are comparatively better approaches for understanding complex systems. For instance we observed that GLE can reproduce the underlying dynamics of Niño-3 index with higher accuracy than LE.

In the second part, using the knowledge gained from the first part, we employed KM to study the underlying mechanism of paleoclimate proxy records (i.e., δ^{18}O and dust concentrations) obtained from the NGRIP ice core. The central point here was to

examine the stability configuration of the coupled δ^{18}O–dust process by reconstructing its potential landscape. We observed that the minimums' positions in the monostable potential of the δ^{18}O are controlled by the value of the dust record. According to this finding, we could confirm the existence of possible couplings between the δ^{18}O and dust time series. Additionally, we found non-vanishing higher-order KM coefficients for δ^{18}O, indicating the presence of discontinuities in the record. This led to the conclusion that the conventional LE is not suitable to fully describe the underlying process of the δ^{18}O and requires adding a new term that takes into account jumps. Conversely, we observed that isolated dust record is a continuous process that LE can describe it successfully.

6.2.1. Outlook

Understanding the physical causes of past abrupt climate changes is crucial for improving the predictive capability of Earth System Models. In chapter 4, we considered only coupling between two Greenland ice core proxies (the δ^{18}O and dust records) to understand the triggering mechanism of abrupt transitions. However, it has been argued that abrupt transitions can be the results of a more complex interplay of the North Atlantic and Nordic Sea's ice cover and several climatic subsystems, such as ice sheets or the East Asian Monsoon system [32]. As an extension of the approach described in this chapter, we suggest analogous analyses for other pairs of Greenland proxies. Additionally, we considered classical Gaussian random noise to describe the high-frequency variability. However, it has been shown that selecting stochastic forcing other than Gaussian can provide insight into the abrupt climate changes [223]. Therefore, in real-world phenomena, it is vital to consider an adaptive SDE framework capable of reconstruct the underlying dynamics of systems driven by, e.g., Lévy noise. The adaptability of the three investigated SDE models can be further investigated for reproducing dynamical characteristics of multi-variate time series.

6.3. Conclusion from study 4 & 5: Predicting Complex Systems Dynamics Using Artificial Neural Network

In the last part of this thesis (chapter 5), we focused on developing an ANN algorithm that can predict climate variability from limited-time series with a low signal-to-noise ratio. For this purpose, we chose Echo State Network (ESN), which is suitable for investigating complex nonlinear time series. The challenging part of predicting climate time series was understanding the dynamics of high-frequency variability that can influence the prediction of low-frequency variability. To tackle this problem, we estimated the future behavior of high-frequency variability from its past history using the PNF method [30]. By providing the essential information about the high-frequency variability of the target system (e.g., external forcings), we could predict different climate indices, such as ENSO, PDO, and AMO. The rationale behind our proposed model (called ESN-PNF) was that climate can be approximately decomposed into slow and fast variability modes, where the slow mode is only mildly perturbed by high-frequency forcing. Our results demonstrated that the ESN-PNF model can capture the key features of the target indices and predict their dynamics with a reasonable long forecast horizon. For instance, we could achieve long-term perdition of ENSO indices far beyond the spring barrier. The improved predictability demonstrated that interactions across multiple time scale play a crucial role in generating the dynamics of ENSO, PDO, and AMO indices. Moreover, in contrast to different ANNs and deep-learning models [72], we showed that our model does not require massive data for training. This

feature is especially important considering that additional data provided by process-based models might cause unintended biases and structural errors.

6.3.1. Outlook

In chapter 5, we focused on predicting complex systems' behavior using the ANN approach. As an extension of this methodology, one can combine physics-based models with different ANN algorithms to enhance the predictions of the system in question. Even though physics-based models advanced our understanding of natural principles, they mostly rely on existing knowledge and cannot extract more information from available data. On the other hand, even though ANNs are skillful in predicting without theoretical knowledge of the underlying dynamics, their large data requirements make the training procedure computationally expensive. Moreover, the performance of ANNs can be adversely affected when the quality of data is low. Understandably, there is a consensus [101, 151, 176, 186] that ANNs can be assisted by providing additional information derived from mathematical models and physical laws. This prior knowledge about the system can constrain the space of possible solutions by neglecting irrelevant ones, which results in faster convergence to the optimal solution and better predictive performance.

Appendices

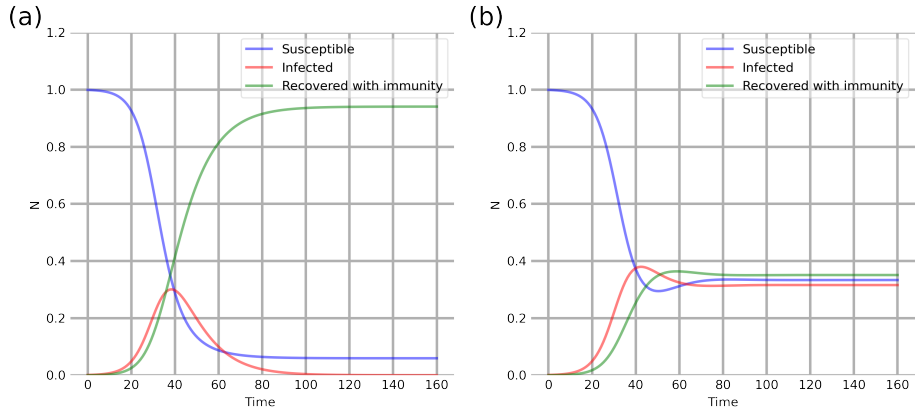

FIGURE A.1.: Schematic illustration of the epidemic spreading models. Panels **(a)** and **(b)** show a single realization of the SIR and SIRS models respectively with $\beta = 0.3$, $\gamma = 0.1$ and $\delta = 0.1$.

A. Using Event Synchrony Measures for Network Inference From the Timing of Events

A.1. SIRS Model

The Susceptible–Infected–Recovered (SIR) model emerged from the area of epidemiology to understand the time evolution of infectious diseases in a fixed population of N. This model assumes that the interaction between individuals can be descried by two specific parameters, the disease transmission and removal rates. However, in various real-world examples like the seasonal influenza, in which an infection can be spread repeatedly among a given population, it is not realistic to presume individuals have permanent immunity . In this thesis, we therefore employ one variant of the well-known SIRS (Susceptible-Infected-Recovered-Susceptible) model of epidemic spreading [83]. This model comprises three distinctive "health" states of each individual: susceptible (S), infected (I) and recovered (R). In the conventional SIRS model, the densities of individuals in each of those three states follow the macroscopic evolutionary equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS + \delta R, \\ \frac{dI}{dt} &= \beta IS - \gamma I, \\ \frac{dR}{dt} &= \gamma I - \delta R, \end{aligned} \tag{A.1}$$

where β and γ denote the infection rate and natural recovery rate, respectively, while δ describes the probability of recovered individuals to lose their immunity.

A.2. Receiver operating characteristic (ROC)

Instead of predicting the class of binary events, we can alternatively predict the probabilities of the classes. This alternate approach enables us to choose the optimal threshold in which the model exhibit better performance in classification task. The receiver operating characteristic (ROC) [53, 73] is ubiquitously employed in many areas such as medicine, natural hazards, and machine learning, when predicting the probability of binary classification problems. The ROC curve shows the trade-off between true positive rates (TPR) and false positive rates (FPR) over a range of different cut-off points of a parameter. The TPR rate (also known as sensitivity rate) indicates the proportion of positive classes that the model could correctly classify, and by contrast, FPR measures the proportion of negative classes that have been incorrectly classified in the presence of a predetermined condition. The area under the ROC curve (AUC) provides a convenient way to summarize the model skill.This value reflects the overall accuracy of a classification approach and demonstrate how a model can distinguish between different classes. A ROC curve that falls together with the diagonal (TPR = FPR) is characterized by a value of AUC = 0.5, which indicates a classification with a performance equal to that of a random forecast (and, hence, the inability to discriminate between true and false links), while AUC = 1 would indicate a perfect classification. The probabilistic interpretation of the AUC can be demonstrated as follows.

$$AUC = P(p(X') :> p(X)|L(X') = 1, L(X) = 0) \tag{A.2}$$

Where p(X) denotes the probability score that the model assigns to a random variable X if it belongs to the binary class 1, and the L(.) determines class labels. According to the Eq. A.2, AUC shows that the model give higher score to X' rather than X.

B. Fundamentals of Dynamical Systems

B.1. Stochastic calculus

Equation (2.15) with a given initial condition $x(t_0) = x_0$ has a unique solution which satisfies the following integral form:

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x(t'))dt' + \int_{t_0}^{t} g(x(t'))\eta(t')dt' \tag{B.1}$$

A Wiener process is non-smooth and nowhere differentiable, hence translating the second integral in the equation (B.1) using a conventional Riemann sum is not uniquely defined. To interpret the noise term, two different formulations of stochastic calculus have been introduced for computing the solutions of SDEs; known as Itô and Stratonovich calculi, respectively [103, 199].

In the Itô prescription, the evaluations of the function $g(x(t))$ are uncorrelated with the (infinitesimal) increments of the Wiener process $dW(t)$. In fact, the Itô integral is defined as the limit of a left Riemann sum (where the function g(x(t)) is evaluated at the left of the interval $[t, t + \Delta t]$) with an Itô correction. The resulting stochastic integration in the Itô scheme is given:

$$\int_{t_0}^{t} W(s)dW(s)ds = \frac{1}{2}[(W^2(t) - W^2(t_0)) - (t - t_0)]\,. \tag{B.2}$$

Although the Itô convention does not preserve the chain rule of calculus, employing Itô's Lemma maintains the Martingale property. Owing to this property, the Itô calculus is used extensively in finance [191].

The most common alternative to the Itô integral that does satisfy the chain rule of classical calculus is the Stratonovich scheme. From that point of view, a function can be evaluated at the midpoint of the time interval $[t, t + \Delta t]$. Because the midpoint selection rule is associated with the finite noise autocorrelation [153], the Martingale property does not hold. In contrast to the Itô, the Stratonovich calculus approximates the Wiener process as the limit of a correlated process when the correlation time approaches zero:

$$\int_{t_0}^{t} W(s) \circ dW(s) = \frac{1}{2}(W^2(t) - W^2(t_0)) \tag{B.3}$$

This approximation leads to difficulties e.g. for the calculation of expectation values, since stochastic variables and noise are not independent: $< x(t)\eta(t) > \neq 0$. It should be underlined that the Itô and Stratonovich calculi have the same solution if their drift terms fulfill the following relationship, which is called Itô-Stratonovich drift correction:

$$f_S(x(t)) = f_I(x(t)) - \frac{1}{2}(g(x(t))\frac{\partial g(x(t))}{\partial x}\,, \tag{B.4}$$

where f_I denotes the drift of the Itô calculus and f_S the drift of the Stratonovich calculus.

Even though both interpretations are mathematically consistent, one obvious question that may arise is which interpretation is the right one for desribing or approximating a particular set of physical processes. In order to answer this, we have to

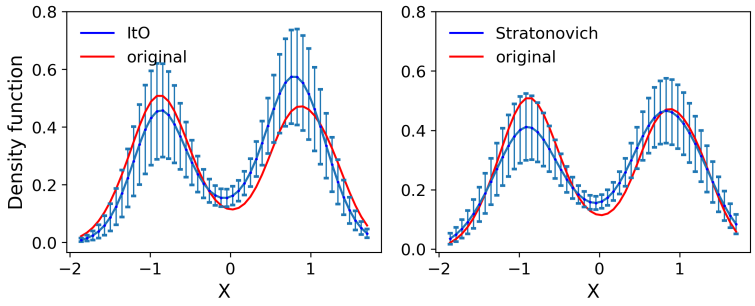

FIGURE B.1.: Comparison between PDFs obtained from the Itô scheme (left panel) and the Stratonovich scheme (right panel) for the Ca^{2+} ice core time series.

look at the origin of the noise in the stochastic system. It has been shown that, if the relaxation time of a system is large enough in comparison with the noise correlation time, then the Itô interpretation is appropriate. On the other hand, if the noise is colored, i.e., it has finite correlation time, the limiting SDE must be treated in the Stratonovich framework [88, 112, 153].

As noted above, different stochastic calculi (Itô or Stratonovich) are associated with different kinds of discritization for numerical integration [183]. It must hence be stated upfront which stochastic calculus is going to be considered. The simplest and most widely used discretization scheme to numerically integrate SDEs is the Euler-Maruyama method, which converges to the Itô interpretation:

$$x(t+\Delta t) = x(t) + f(x(t))\Delta t + g(x(t))\eta(t)\sqrt{\Delta t} \tag{B.5}$$

Another numerical method that we used in this paper (for approximating the Ca^{2+} ice core time series), called Heun method[68], leads to the Stratonovich scheme. This method is an example of a predictor-corrector method in which the predictor is calculated by a simple Euler type integration as follows:

$$\hat{x}(t+\Delta t) = x(t) + f(x(t))\Delta t + g(x(t))\sqrt{\Delta t}\eta(t) \tag{B.6}$$

$$\begin{aligned} x(t+\Delta t) = x(t) &+ \frac{1}{2}(f(x(t)) + f(\hat{x}(t+\Delta t))\Delta t \\ &+ \frac{1}{2}(g(x(t)) + g(\hat{x}(t+\Delta t))\sqrt{\Delta t}\eta(t) \end{aligned} \tag{B.7}$$

In this thesis, based on different systems we utilized both methods.

B.1.1. Comparison between Itô and Stratonovich

We compared the performances of both stochastic calculi and corresponding discretization schemes (Euler-Maruyama and Heun, respectively) for the Ca^{2+} ice core time series. According to Fig. (B.1), we observed that the Stratonovich calculus performs better than the Itô calculus in terms of approximating the PDF.

B.2. Adjoint Fokker-Planck Equation (AFPE)

The estimation of KM coefficients obtained from equation 2.17 requires small $\tau \to 0$, whereas we always deal with time series that are recorded at finite sampling intervals [114, 117]. Therefore the effect of finite-time distortions can reduce the accuracy of the estimated drift and diffusion terms using KM method. An elegant way to obtain

Time Series	KM	MZ	EMR
OU additive noise	0.00003	0.00032	0.00003
OU multiplicative noise	0.0013	0.0021	0.004
OU colored noise	0.002	0.00001	0.00001
S&P500	0.05	0.03	0.06
Niño-3	0.01	0.001	0.001
DW additive noise	0.0004	0.00042	0.0021
DW multiplicative noise	0.0008	0.0009	0.001
$[Ca^{2+}]$	0.004	0.009	0.009

TABLE B.1.: Summary of Mean Squared Error (MSE) between PDFs of different systems and corresponding averaged PDF of simulated time series over several realizations.

the effect of finite-time distortions is adjoint Fokker-Planck equation (AFPE) [82]. If we define the FPE operator as $\mathcal{J}$,

$$\mathcal{J} = \frac{\partial}{\partial x} D_1 + \frac{\partial^2}{\partial x^2} D_2 \tag{B.8}$$

Then the solution of FPE with the initial condition $\delta(x' - x)$ can be defined as follows:

$$p(x', t + \tau | x, t) = \exp \mathcal{J}(x') \tau \delta(x' - x) \tag{B.9}$$

By substituting the $p(x', t + \tau | x, t)$ in the $D_j(x) = \frac{1}{j!} \int_{-\infty}^{\infty} (x(t + \tau) - x(t))^j p(x(t + \tau) | x(t)), dx$, which is the continuous version of Eq. 2.19, we obtain:

$$D^m(x) = \exp \mathcal{J}^+(x') \tau (x' - x)^j | x' = x \tag{B.10}$$

Where $\mathcal{J}^+$ denotes the adjoint FPE operator. Using Heisenberg approach, Lade[114] proposed that the solution of the following partial differential equation will give us $D^m(x)$.

$$\frac{\partial}{\partial t} Q_j(x, t) = \mathcal{J}^+ Q_j(x, t) \tag{B.11}$$

with initial condition of $Q_j(x, t) = \frac{1}{j!}(x' - x)^j$.

B.3. Supplementary Figures

Time Series	KM	MZ	EMR
OU additive noise	0.00013	0.00021	0.00034
OU multiplicative noise	0.00038	0.0003	0.00027
OU colored noise	0.07	0.00004	0.00013
S&P500	0.002	0.001	0.0008
Niño-3	0.01	0.003	0.006
DW additive noise	0.00015	0.00002	0.00031
DW multiplicative noise	0.0009	0.005	0.001
$[Ca^{2+}]$	0.026	0.004	0.004

TABLE B.2.: The table illustrates MSE (as an error metric) between ACFs of original systems and corresponding average ACFs of simulated time series.

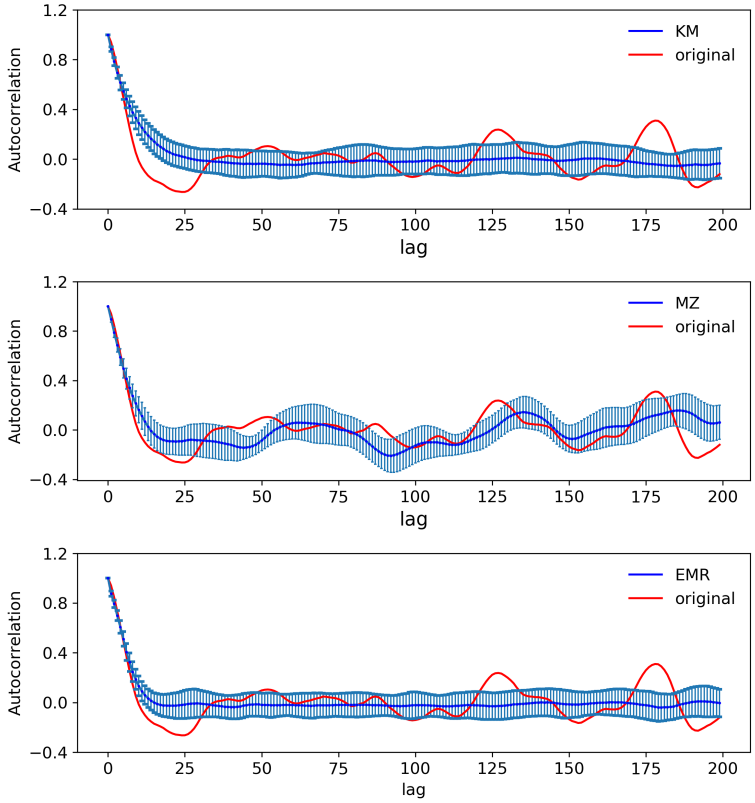

FIGURE B.2.: Comparison between ACFs obtained from calibrating different models (KM,MZ,EMR) on the first half of Niño-3 monthly SST.

C. Predicting Climate Variability Using Machine learning Approach

C.1. Forecasting skill scores

Extreme events are rare events that ubiquitously observed in many natural systems from financial crisis and natural disaster. Therefore, accurately forecasting rare high-impact events, like extreme temperature and high precipitation, will be of high value for decision makers. The usage of binary classification happens frequently when we need to characterize and predict extreme events in the observed time series. In such models, extremes are often displayed as probability levels such as above or below a critical value. Several scores have been developed to evaluate the predictive capability of binary classifiers for choosing "the best" model that can identifies these rare events. In this study we employ Heidke Skill Score and probability of detection.

C.1.1. Heidke Skill Score (HSS)

One of the scores that is widely applied for assessing categorical forecast performance is the Heidke Skill Score (HSS) [13], known outside of meteorology as kappa. The derivation of HSS is based on information summarized in a contingency table, known as an error matrix. This table indicates the relationship between the forecasts and the respective observation.

	Observation = Yes	Observation = No
Forecasting = Yes	TP	FP
Forecasting = No	FN	TN

The HSS measures the accuracy of a forecast with respect to a randomly generated forecast, adjusted to predictions that are correct by chance:

$$HSS = \frac{TP + TN - CRF}{N - CRF} \tag{C.1}$$

where TP and TN stand for true positives and true negatives, respectively, N is the total number of possible events, and CRF indicates the number of correct random forecasts, which can be calculated as follows:

$$CRF = \frac{(TP + FN)(TP + FP) + (TN + FN)(TN + FP)}{N} . \tag{C.2}$$

Negative HSS values imply that the forecast skill of the model is worse than a random forecast, HSS = 0 indicates that the forecast is just as good as the random forecast, and HSS = 1 would indicate a perfect binary forecast.

C.1.2. Probability of detection (POD)

The probability of detection is another quantitative evaluation metric widely used to interpret the model's success in forecasting events correctly. POD measures the number of correct event forecasts divided by the total number of observed events.

$$POD = \frac{TP}{TP + FN} \tag{C.3}$$

C.2. Supplementary Figures

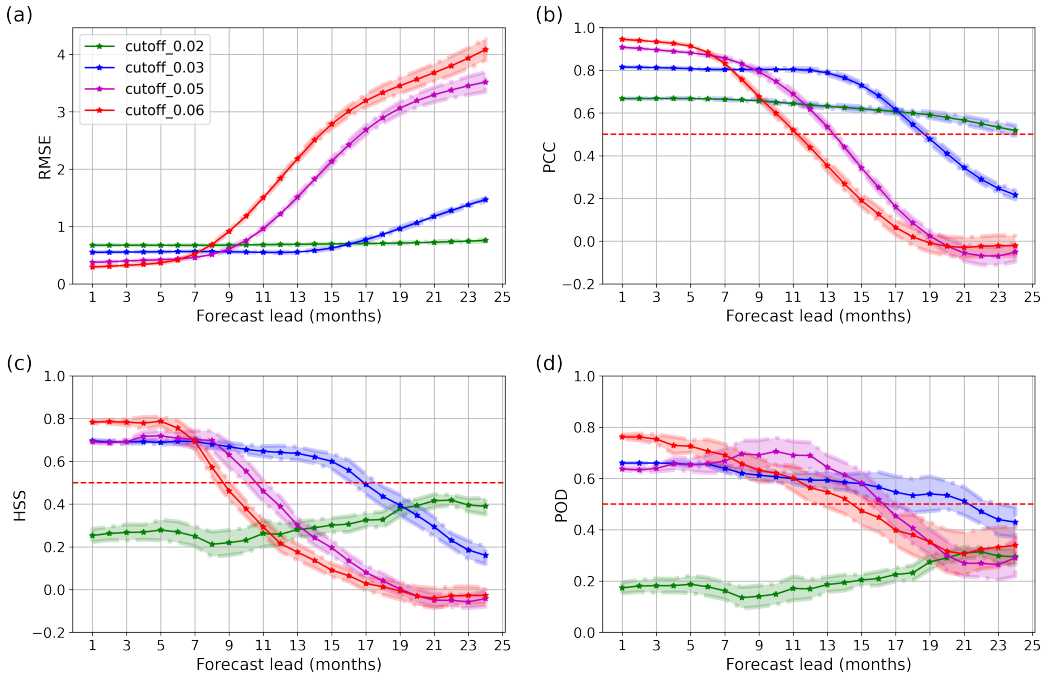

FIGURE C.1.: Summary of the ESN-PNF forecast skill for Niño-3 index for different cutoff choices to decompose the index into slow and fast components. Panels **(a)** and **(b)** show the prediction skill of the model in terms of Root Mean Square Error (RMSE) and Pearson correlation coefficient (PCC) between predicted and observed values. The ability of the model to detect El Niño events (i.e., months with Niño-3 index above 1 $^{\circ}C$) is assessed by two binary classifiers, the Heidke score skill (HSS) and the Probability of detection (POD), panels **(c)** and **(d)**, respectively.The model is trained with a fixed training data length $T = 1092$ and the prediction task starts from 1982. Here the solid lines indicate the average over 100 different realizations and colored shadings around lines correspond to $\pm 1\sigma$. Different colors, as indicated in the legend of panel **(a)**, correspond to results for different cutoff thresholds to decompose the ENSO index. In order to determine the optimal cutoff value, we additionally investigated the dependency of the binary forecast skill of predicting El Niño events, using HSS and POD metrics. The ESN model exhibits overall best performance at a cutoff value of $C = 0.03$ (months^{-1}). Even though a cutoff at $C = 0.02$ (months^{-1}) displays a higher correlation skill for very long lead times, the forecast of El Niño events, and also the reproduction of statistical properties such as the ACF and PDF is worse in this case.

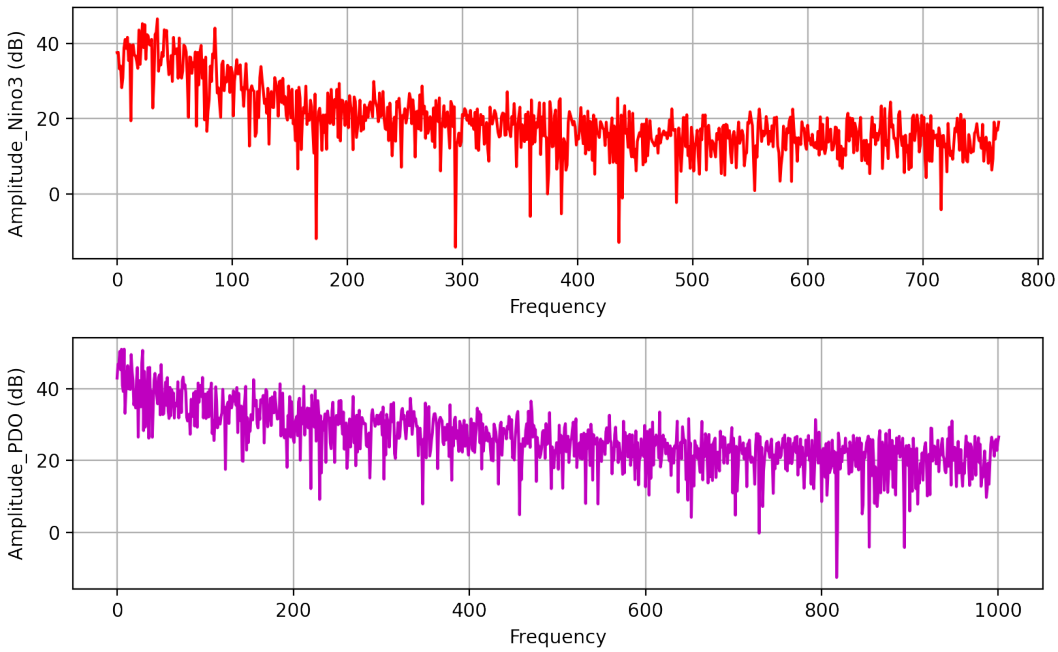

FIGURE C.2.: The frequency spectrum of Niño-3 Aand PDO

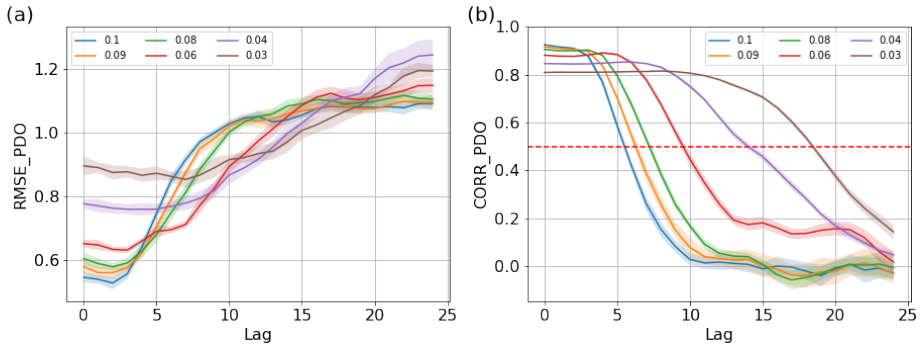

FIGURE C.3.: Summary of the ESN forecast skill for PDO index at different cutoff choices used to decompose the index into slow and fast components. Panels **(a)** and **(b)** show the prediction skill of the model in terms of Root Mean Square Error (RMSE) and Pearson correlation coefficient (PCC) between predicted and observed values

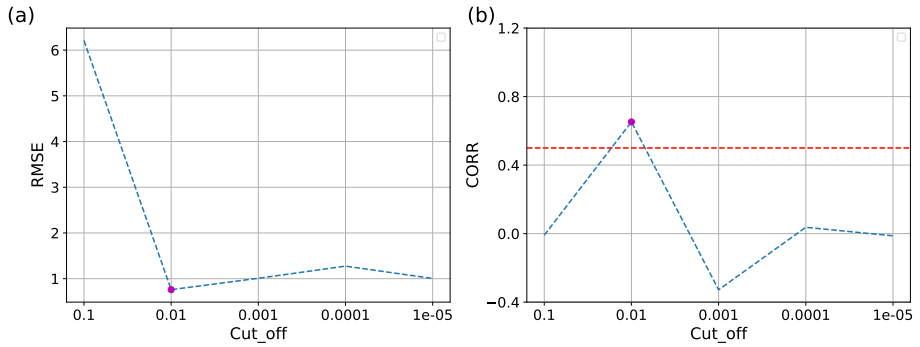

FIGURE C.4.: Summary of the ESN forecast skill for PDO index for different cutoff choices to decompose the index into slow and fast components. This figure illustrates the boundary - marked in magenta- in which we can narrow down our search for the best low-pass filter cutoff.

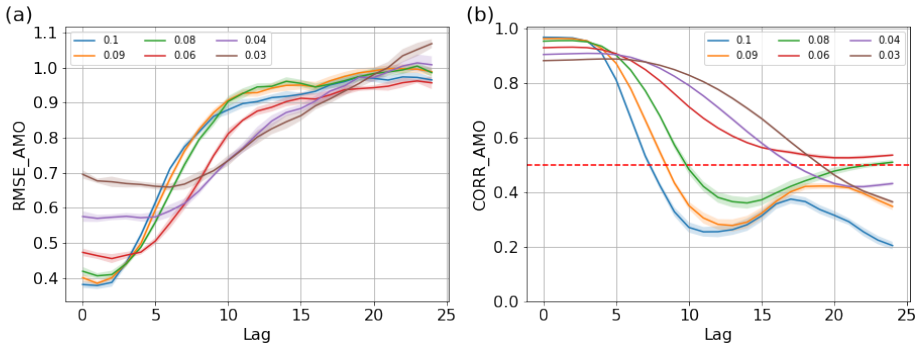

FIGURE C.5.: Summary of the ESN forecast skill for AMO index at different cutoff choices used to decompose the index into slow and fast components. Panels **(a)** and **(b)** show the prediction skill of the model in terms of Root Mean Square Error (RMSE) and Pearson correlation coefficient (PCC) between predicted and observed values

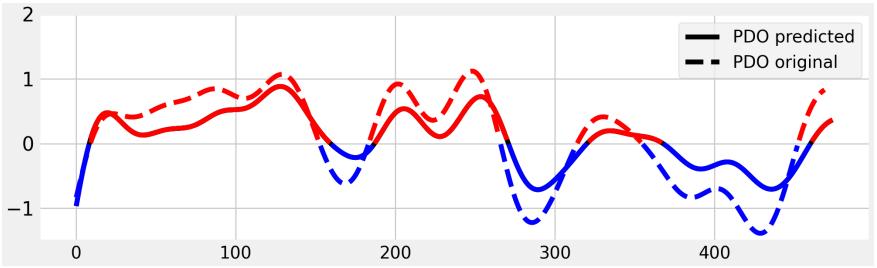

FIGURE C.6.: Detecting phase changes from five years moving average of predicted PDO index (solid lines). As it can be seen the ESN-PNF model can perfectly detect shifts between warm and cold phases at the 18-month lead-time.

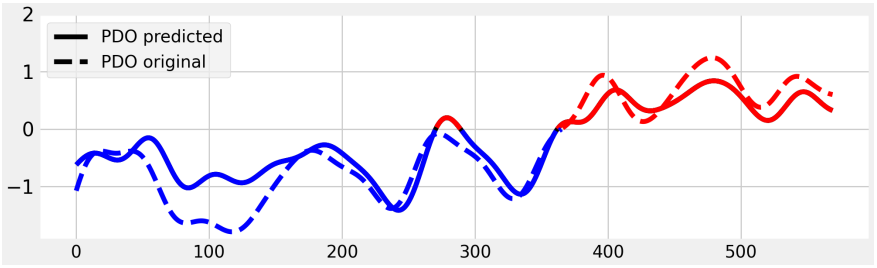

FIGURE C.7.: Detecting phase changes from five years moving average of predicted AMO index (solid lines). As it can be seen the ESN-PNF model can successfully detect shifts between warm and cold phases at the 16-month lead-time.

Bibliography

[1] Amir AghaKouchak, David Easterling, Kuolin Hsu, Siegfried Schubert, and Soroosh Sorooshian. *Extremes in a changing climate: detection, analysis and uncertainty*. Vol. 65. Springer Science & Business Media, 2012.

[2] Michael A Alexander, Ileana Bladé, Matthew Newman, John R Lanzante, Ngar-Cheung Lau, and James D Scott. „The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans“. In: *Journal of Climate* 15.16 (2002), pp. 2205–2231.

[3] Michael A Alexander, Ludmila Matrosova, Cécile Penland, James D Scott, and Ping Chang. „Forecasting Pacific SSTs: Linear inverse model predictions of the PDO“. In: *Journal of Climate* 21.2 (2008), pp. 385–402.

[4] Jürgen Alheit, Priscilla Licandro, Steve Coombs, Alberto Garcia, Ana Giráldez, Maria Teresa Garcia Santamarıa, Aril Slotte, and Athanassios C Tsikliras. „Reprint of “Atlantic Multidecadal Oscillation (AMO) modulates dynamics of small pelagic fishes and ecosystem regime shifts in the eastern North and Central Atlantic”“. In: *Journal of Marine Systems* 133 (2014), pp. 88–102.

[5] Mohamed Alloghani, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, and Ahmed J Aljaaf. „A systematic review on supervised and unsupervised machine learning algorithms for data science“. In: *Supervised and unsupervised learning for data science* (2020), pp. 3–21.

[6] Jose Alvarez-Ramirez, Eduardo Rodriguez, and Juan Carlos Echeverrıa. „Detrending fluctuation analysis based on moving average filtering“. In: *Physica A: statistical mechanics and its applications* 354 (2005), pp. 199–219.

[7] Soon-Il An and Fei-Fei Jin. „Nonlinearity and asymmetry of ENSO“. In: *Journal of Climate* 17.12 (2004), pp. 2399–2412.

[8] M Anvari, B Werther, G Lohmann, M Wächter, J Peinke, and H-P Beck. „Suppressing power output fluctuations of photovoltaic power plants“. In: *Solar Energy* 157 (2017), pp. 735–743.

[9] Rubén Banderas, Jorge Alvarez-Solas, Alexander Robinson, and Marisa Montoya. „An interhemispheric mechanism for glacial abrupt climate change“. In: *Climate Dynamics* 44.9 (2015), pp. 2897–2908.

[10] Yaneer Bar-Yam. *Dynamics of complex systems*. Addison-Wesley Reading, MA, 1997.

[11] Yaneer Bar-Yam, Susan R McKay, and Wolfgang Christian. „Dynamics of complex systems (Studies in nonlinearity)“. In: *Computers in Physics* 12.4 (1998), pp. 335–336.

[12] Albert-László Barabási and Réka Albert. „Emergence of Scaling in Random Networks“. In: *Science* 286.5439 (1999), pp. 509–512.

[13] Anthony G Barnston. „Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score“. In: *Weather and Forecasting* 7.4 (1992), pp. 699–709.

[14] George Bebis and Michael Georgiopoulos. „Feed-forward neural networks“. In: *IEEE Potentials* 13.4 (1994), pp. 27–31.

[15] Debarun Bhattacharjya, Dharmashankar Subramanian, and Tian Gao. „Proximal Graphical Event Models“. In: *Proceedings of the 32Nd International Conference on Neural Information Processing Systems*. NIPS'18. Montréal, Canada: Curran Associates Inc., 2018, pp. 8147–8156.

[16] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. „Complex networks: Structure and dynamics“. In: *Physics Reports* 424.4-5 (2006), pp. 175–308.

[17] Niklas Boers, Bodo Bookhagen, Henrique MJ Barbosa, Norbert Marwan, Jürgen Kurths, and JA Marengo. „Prediction of extreme floods in the eastern Central Andes based on a complex networks approach“. In: *Nature Communications* 5 (2014), p. 5199.

[18] Niklas Boers, Bodo Bookhagen, Norbert Marwan, Jürgen Kurths, and José Marengo. „Complex networks identify spatial patterns of extreme rainfall events of the South American Monsoon System“. In: *Geophysical Research Letters* 40.16 (2013), pp. 4386–4392.

[19] Niklas Boers, Mickael D Chekroun, Honghu Liu, Dmitri Kondrashov, Denis-Didier Rousseau, Anders Svensson, Matthias Bigler, and Michael Ghil. „Inverse stochastic–dynamic models for high-resolution Greenland ice core records“. In: (2017).

[20] Niklas Boers, Michael Ghil, and Denis-Didier Rousseau. „Ocean circulation, ice shelf, and sea ice interactions explain Dansgaard–Oeschger cycles“. In: *Proceedings of the National Academy of Sciences* 115.47 (2018), E11005–E11014.

[21] Niklas Boers, Jürgen Kurths, and Norbert Marwan. „Complex systems approaches for Earth system data analysis“. In: *Journal of Physics: Complexity* 2.1 (2021), p. 011001.

[22] Niklas Boers, Aljoscha Rheinwalt, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths. „A complex network approach to investigate the spatiotemporal co-variability of extreme rainfall“. In: *Machine Learning and Data Mining Approaches to Climate Science*. Springer, 2015, pp. 163–174.

[23] Lorenzo Boninsegna, Feliks Nüske, and Cecilia Clementi. „Sparse learning of stochastic dynamical equations“. In: *The Journal of chemical physics* 148.24 (2018), p. 241723.

[24] Dirk Brockmann and Dirk Helbing. „The hidden geometry of complex, network-driven contagion phenomena“. In: *science* 342.6164 (2013), pp. 1337–1342.

[25] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. „Discovering governing equations from data by sparse identification of nonlinear dynamical systems“. In: *Proceedings of the national academy of sciences* 113.15 (2016), pp. 3932–3937.

[26] Enrique Castillo. „Functional Networks“. In: *Neural Processing Letters* 7.3 (1998), pp. 151–159.

[27] Chris Chatfield. *Time-series forecasting*. Chapman and Hall/CRC, 2000.

[28] Ashesh Chattopadhyay, Pedram Hassanzadeh, and Devika Subramanian. „Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network“. In: *Nonlinear Processes in Geophysics* 27.3 (2020), pp. 373–389.

[29] Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. „Analog forecasting of extreme-causing weather patterns using deep learning“. In: *Journal of Advances in Modeling Earth Systems* 12.2 (2020), e2019MS001958.

[30] Mickaël David Chekroun, Dmitri Kondrashov, and Michael Ghil. „Predicting stochastic systems by noise sampling, and application to the El Niño-Southern Oscillation“. In: *Proceedings of the National Academy of Sciences* 108.29 (2011), pp. 11766–11771.

[31] Su Chen. „Optimal bandwidth selection for kernel density functionals estimation“. In: *Journal of Probability and Statistics* 2015 (2015).

[32] Hai Cheng, Ashish Sinha, Francisco W. Cruz, Xianfeng Wang, R. Lawrence Edwards, Fernando M. D'Horta, Camila C. Ribas, Mathias Vuille, Lowell D. Stott, and Augusto S. Auler. „Climate change patterns in Amazonia and biodiversity“. In: *Nature Communications* 4 (2013), p. 1411.

[33] Yoshimitsu Chikamoto, Masahide Kimoto, Masayoshi Ishii, Takashi Mochizuki, Takashi T Sakamoto, Hiroaki Tatebe, Yoshiki Komuro, Masahiro Watanabe, Toru Nozawa, Hideo Shiogama, et al. „An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC“. In: *Climate Dynamics* 40.5-6 (2013), pp. 1201–1222.

[34] Alexandre J Chorin, Ole H Hald, and Raz Kupferman. „Optimal prediction with memory“. In: *Physica D: Nonlinear Phenomena* 166.3-4 (2002), pp. 239–257.

[35] Peter U Clark, Shawn J Marshall, Garry KC Clarke, Steven W Hostetler, Joseph M Licciardi, and James T Teller. „Freshwater forcing of abrupt climate change during the last glaciation“. In: *Science* 293.5528 (2001), pp. 283–287.

[36] William D Collins, Cecilia M Bitz, Maurice L Blackmon, Gordon B Bonan, Christopher S Bretherton, James A Carton, Ping Chang, Scott C Doney, James J Hack, Thomas B Henderson, et al. „The community climate system model version 3 (CCSM3)“. In: *Journal of Climate* 19.11 (2006), pp. 2122–2143.

[37] Louise K Comfort. „Self-organization in complex systems“. In: *Journal of Public Administration Research and Theory: J-PART* 4.3 (1994), pp. 393–410.

[38] Eric Darve, Jose Solomon, and Amirali Kia. „Computing generalized Langevin equations and generalized Fokker–Planck equations“. In: *Proceedings of the National Academy of Sciences* 106.27 (2009), pp. 10884–10889.

[39] Thomas L Delworth and Michael E Mann. „Observed and simulated multidecadal variability in the Northern Hemisphere“. In: *Climate Dynamics* 16.9 (2000), pp. 661–676.

[40] Clara Deser, Marika Holland, Gilles Reverdin, and Michael Timlin. „Decadal variations in Labrador Sea ice cover and North Atlantic sea surface temperatures“. In: *Journal of Geophysical Research: Oceans* 107.C5 (2002), pp. 3–1.

[41] PD Ditlevsen, KK Andersen, and A Svensson. „The DO-climate events are noise induced: statistical investigation of the claimed 1470 years cycle“. In: *Climate of the Past Discussions* 2.6 (2006), pp. 1277–1292.

[42] Peter D Ditlevsen, Katrine Krogh Andersen, and Anders Svensson. „The DO-climate events are probably noise induced: statistical investigation of the claimed 1470 years cycle“. In: *Climate of the Past* 3.1 (2007), pp. 129–134.

[43] Peter D Ditlevsen, Mikkel S Kristensen, and Katrine K Andersen. „The recurrence time of Dansgaard–Oeschger events and limits on the possible periodic component“. In: *Journal of Climate* 18.14 (2005), pp. 2594–2603.

[44] Jonathan F Donges, C-F Schleussner, Jonatan F Siegmund, and Reik V Donner. „Event coincidence analysis for uantifying statistical interrelationships between event time series“. In: *The European Physical Journal Special Topics* 225.3 (2016), pp. 471–487.

[45] Jonathan F Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths. „Complex networks in climate dynamics“. In: *The European Physical Journal Special Topics* 174.1 (2009), pp. 157–179.

[46] Reik Donner, Susana Barbosa, Jürgen Kurths, and Norbert Marwan. „Understanding the Earth as a Complex System–recent advances in data analysis and modelling in Earth sciences“. In: *The European Physical Journal Special Topics* 174.1 (2009), pp. 1–9.

[47] Reik V. Donner, Marc Wiedermann, and Jonathan F. Donges. „Complex Network Techniques for Climatological Data Analysis“. In: *Nonlinear and Stochastic Climate Dynamics*. Ed. by Christian L. E. Franzke and Terence J.Editors O'Kane. Cambridge University Press, 2017, pp. 159–183.

[48] Uwe Einmahl and David M Mason. „Uniform in bandwidth consistency of kernel-type function estimators“. In: *The Annals of Statistics* 33.3 (2005), pp. 1380–1403.

[49] Nicole El Karoui, Shige Peng, and Marie Claire Quenez. „Backward stochastic differential equations in finance“. In: *Mathematical finance* 7.1 (1997), pp. 1–71.

[50] David B Enfield, Alberto M Mestas-Nuñez, and Paul J Trimble. „The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US“. In: *Geophysical Research Letters* 28.10 (2001), pp. 2077–2080.

[51] Ernesto Estrada. *The structure of complex networks: theory and applications*. Oxford University Press, 2012.

[52] F Farahpour, Z Eskandari, A Bahraminasab, GR Jafari, F Ghasemi, Muhammad Sahimi, and M Reza Rahimi Tabar. „A Langevin equation for the rates of currency exchange based on the Markov analysis“. In: *Physica A: Statistical Mechanics and its Applications* 385.2 (2007), pp. 601–608.

[53] Tom Fawcett. „An introduction to ROC analysis“. In: *Pattern Recognition Letters* 27.8 (2006), pp. 861–874.

[54] Thomas Finley and Thorsten Joachims. „Supervised clustering with support vector machines“. In: *Proceedings of the 22nd international conference on Machine learning*. 2005, pp. 217–224.

[55] Chris K Folland, Andrew W Colman, David P Rowell, and Mike K Davey. „Predictability of northeast Brazil rainfall and real-time forecast skill, 1987–98“. In: *Journal of Climate* 14.9 (2001), pp. 1937–1958.

[56] Andrew Cadle Fowler, Anna C Fowler, and AC Fowler. *Mathematical models in the applied sciences*. Vol. 17. Cambridge University Press, 1997.

[57] Rudolf Friedrich and Joachim Peinke. „Description of a turbulent cascade by a Fokker-Planck equation“. In: *Physical Review Letters* 78.5 (1997), p. 863.

[58] Rudolf Friedrich and Joachim Peinke. „Statistical properties of a turbulent cascade“. In: *Physica D: Nonlinear Phenomena* 102.1-2 (1997), pp. 147–155.

[59] Rudolf Friedrich, Joachim Peinke, Muhammad Sahimi, and M Reza Rahimi Tabar. „Approaching complexity by stochastic methods: From biological systems to turbulence“. In: *Physics Reports* 506.5 (2011), pp. 87–162.

[60] GCOS-AOPC/OOPC. *Working group on surface pressure*. `https://psl.noaa.gov/gcos_wgsp/Timeseries/Nino3/`. (accessed December 15, 2019).

[61] Fatemeh Ghasemi, Muhammad Sahimi, J Peinke, R Friedrich, G Reza Jafari, and M Reza Rahimi Tabar. „Markov analysis and Kramers-Moyal expansion of nonstationary stochastic processes with application to the fluctuations in the oil price“. In: *Physical Review E* 75.6 (2007), p. 060102.

[62] Victor Gomez. „The use of Butterworth filters for trend and cycle estimation in economic time series“. In: *Journal of Business & Economic Statistics* 19.3 (2001), pp. 365–373.

[63] Parameswaran Gopikrishnan, Vasiliki Plerou, Yan Liu, LA Nunes Amaral, Xavier Gabaix, and H Eugene Stanley. „Scaling and correlation in financial time series“. In: *Physica A: Statistical Mechanics and its Applications* 287.3-4 (2000), pp. 362–373.

[64] Emily M Gordon, Elizabeth A Barnes, and James Wilson Hurrell. „Oceanic harbingers of Pacific Decadal Oscillation predictability in CESM2 detected by neural networks“. In: *Geophysical Research Letters* (2021), e2021GL095392.

[65] Rydin Gorjao, Benjamin Schäfer, Dirk Witthaut, and Christian Beck. „Spatio-temporal complexity of power-grid frequency fluctuations“. In: *New Journal of Physics* 23.7 (2021), p. 073016.

[66] Leonardo Rydin Gorjão, Jan Heysel, Klaus Lehnertz, and M Reza Rahimi Tabar. „Analysis and data-driven reconstruction of bivariate jump-diffusion processes“. In: *Physical Review E* 100.6 (2019), p. 062127.

[67] Alex Graves. „Generating sequences with recurrent neural networks“. In: *arXiv preprint arXiv:1308.0850* (2013).

[68] A Greiner, W Strittmatter, and J Honerkamp. „Numerical integration of stochastic differential equations“. In: *Journal of Statistical Physics* 51.1-2 (1988), pp. 95–108.

[69] Volker Grimm. „Mathematical models and understanding in ecology“. In: *Ecological modelling* 75 (1994), pp. 641–651.

[70] Eric Guilyardi, Andrew Wittenberg, Alexey Fedorov, Mat Collins, Chunzai Wang, Antonietta Capotondi, Geert Jan Van Oldenborgh, and Tim Stockdale. „Understanding El Niño in ocean–atmosphere general circulation models: Progress and challenges“. In: *Bulletin of the American Meteorological Society* 90.3 (2009), pp. 325–340.

[71] Seungwoong Ha and Hawoong Jeong. „Unraveling hidden interactions in complex systems with deep learning“. In: *Scientific reports* 11.1 (2021), pp. 1–13.

[72] Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo. „Deep learning for multi-year ENSO forecasts“. In: *Nature* 573.7775 (2019), pp. 568–572.

[73] James A Hanley and Barbara J McNeil. „[The meaning and use of the area under a receiver operating characteristic (ROC) curve“. In: *Radiology* 143.1 (1982), pp. 29–36.

[74] Hossein Hassani. „Singular spectrum analysis: methodology and comparison“. In: (2007).

[75] Klaus Hasselmann. „Stochastic climate models part I. Theory“. In: *tellus* 28.6 (1976), pp. 473–485.

[76] Rainer Hegger and Gerhard Stock. „Multidimensional Langevin modeling of biomolecular dynamics“. In: *The Journal of chemical physics* 130.3 (2009), p. 034106.

[77] Carmen Hijón, Pep Español, Eric Vanden-Eijnden, and Rafael Delgado-Buscalioni. „Mori–Zwanzig formalism as a practical computational tool“. In: *Faraday discussions* 144 (2010), pp. 301–322.

[78] Jaroslav Hlinka, David Hartman, and Milan Paluš. „Small-world topology of functional connectivity in randomly connected dynamical systems“. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 22.3 (2012), p. 033107.

[79] Jaroslav Hlinka, David Hartman, Martin Vejmelka, Jakob Runge, Norbert Marwan, Jürgen Kurths, and Milan Paluš. „Reliability of inference of directed climate networks using conditional mutual information“. In: *Entropy* 15.6 (2013), pp. 2023–2045.

[80] Arthur E Hoerl and Robert W Kennard. „Ridge regression: applications to nonorthogonal problems“. In: *Technometrics* 12.1 (1970), pp. 69–82.

[81] Christopher J Honey, Rolf Kötter, Michael Breakspear, and Olaf Sporns. „Network structure of cerebral cortex shapes functional connectivity on multiple time scales“. In: *Proceedings of the National Academy of Sciences* 104.24 (2007), pp. 10240–10245.

[82] Christoph Honisch and Rudolf Friedrich. „Estimation of Kramers-Moyal coefficients at low sampling rates“. In: *Physical Review E* 83.6 (2011), p. 066701.

[83] Mevin B Hooten, Jessica Anderson, and Lance A Waller. „Assessing North American influenza dynamics with a statistical SIRS model“. In: *Spatial and Spatio-Temporal Epidemiology* 1.2-3 (2010), pp. 177–185.

[84] Werner Horsthemke. „Noise induced transitions“. In: *Non-Equilibrium Dynamics in Chemical Systems*. Springer, 1984, pp. 150–160.

[85] Qi Hu and Song Feng. „Variation of the North American summer monsoon regimes and the Atlantic multidecadal oscillation“. In: *Journal of Climate* 21.11 (2008), pp. 2371–2383.

[86] Edward L Ince. *Ordinary differential equations*. Courier Corporation, 1956.

[87] Denis J Evans and Gary P Morriss. *Statistical mechanics of nonequilbrium liquids*. ANU Press, 2007.

[88] Kurt Jacobs. *Stochastic processes for physicists: understanding noisy systems*. Cambridge University Press, 2010.

[89] Herbert Jaeger. „Echo state network“. In: *scholarpedia* 2.9 (2007), p. 2330.

[90] Herbert Jaeger, Mantas Lukoševičius, Dan Popovici, and Udo Siewert. „Optimization and applications of echo state networks with leaky-integrator neurons“. In: *Neural networks* 20.3 (2007), pp. 335–352.

[91] Nikola Jajcay, Sergey Kravtsov, George Sugihara, Anastasios A Tsonis, and Milan Paluš. „Synchronization and causality across time scales in El Niño Southern Oscillation“. In: *npj Climate and Atmospheric Science* 1.1 (2018), pp. 1–8.

[92] Mahdi Jalili. „Social power and opinion formation in complex networks“. In: *Physica A: Statistical mechanics and its applications* 392.4 (2013), pp. 959–966.

[93] M Jannesar, A Sadeghi, E Meyer, and GR Jafari. „A Langevin equation that governs the irregular stick-slip nano-scale friction“. In: *Scientific reports* 9.1 (2019), pp. 1–7.

[94] Scot D Johnson, David S Battisti, and ES Sarachik. „Empirically derived Markov models and prediction of tropical Pacific sea surface temperature anomalies“. In: *Journal of climate* 13.1 (2000), pp. 3–17.

[95] Zachary F Johnson, Yoshimitsu Chikamoto, S-Y Simon Wang, Michael J McPhaden, and Takashi Mochizuki. „Pacific decadal oscillation remotely forced by the equatorial Pacific and the Atlantic Oceans“. In: *Climate Dynamics* 55.3 (2020), pp. 789–811.

[96] M Chris Jones, James S Marron, and Simon J Sheather. „A brief survey of bandwidth selection for density estimation“. In: *Journal of the American statistical association* 91.433 (1996), pp. 401–407.

[97] Richard Jordan, David Kinderlehrer, and Felix Otto. „The variational formulation of the Fokker–Planck equation“. In: *SIAM journal on mathematical analysis* 29.1 (1998), pp. 1–17.

[98] Nicolaas G van Kampen. „Remarks on non-Markov processes“. In: *Brazilian Journal of Physics* 28 (1998), pp. 90–96.

[99] Alexey Kaplan, Mark A Cane, Yochanan Kushnir, Amy C Clement, M Benno Blumenthal, and Balaji Rajagopalan. „Analyses of global sea surface temperature 1856–1991“. In: *Journal of Geophysical Research: Oceans* 103.C9 (1998), pp. 18567–18589.

[100] Samuel Karlin and James McGregor. „Ehrenfest urn models“. In: *Journal of Applied Probability* 2.2 (1965), pp. 352–376.

[101] K Kashinath, M Mustafa, A Albert, JL Wu, C Jiang, S Esmaeilzadeh, K Azizzadenesheli, R Wang, A Chattopadhyay, A Singh, et al. „Physics-informed machine learning: case studies for weather and climate modelling“. In: *Philosophical Transactions of the Royal Society A* 379.2194 (2021), p. 20200093.

[102] Ben P Kirtman and Dughong Min. „Multimodel ensemble ENSO prediction with CCSM and CFS“. In: *Monthly Weather Review* 137.9 (2009), pp. 2908–2930.

[103] Peter E Kloeden and Eckhard Platen. „Stratonovich and Itô stochastic taylor expansions“. In: *Mathematische Nachrichten* 151.1 (1991), pp. 33–50.

[104] Stefan Klus, Feliks Nüske, Sebastian Peitz, Jan-Hendrik Niemann, Cecilia Clementi, and Christof Schütte. „Data-driven approximation of the Koopman generator: Model reduction, system identification, and control“. In: *Physica D: Nonlinear Phenomena* 406 (2020), p. 132416.

[105] Mads Faurschou Knudsen, Bo Holm Jacobsen, Marit-Solveig Seidenkrantz, and Jesper Olsen. „Evidence for external forcing of the Atlantic Multidecadal Oscillation since termination of the Little Ice Age“. In: *Nature Communications* 5.1 (2014), pp. 1–8.

[106] D Kondrashov, MD Chekroun, AW Robertson, and M Ghil. „Low-order stochastic model and “past-noise forecasting” of the Madden-Julian Oscillation“. In: *Geophysical Research Letters* 40.19 (2013), pp. 5305–5310.

[107] Dmitri Kondrashov, Mickaël D Chekroun, and Michael Ghil. „Data-driven non-Markovian closure models“. In: *Physica D: Nonlinear Phenomena* 297 (2015), pp. 33–55.

[108] Dmitri Kondrashov, S Kravtsov, Andrew W Robertson, and Michael Ghil. „A hierarchy of data-based ENSO models“. In: *Journal of climate* 18.21 (2005), pp. 4425–4444.

[109] Pavel L Krapivsky, Sidney Redner, and Eli Ben-Naim. *A kinetic view of statistical physics*. Cambridge University Press, 2010.

[110] S Kravtsov, D Kondrashov, and M Ghil. „Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability“. In: *Journal of Climate* 18.21 (2005), pp. 4404–4424.

[111] Rep Kubo. „The fluctuation-dissipation theorem“. In: *Reports on progress in physics* 29.1 (1966), p. 255.

[112] R Kupferman, Grigorios A Pavliotis, and Andrew M Stuart. „Itô versus Stratonovich white-noise limits for systems with inertia and colored multiplicative noise“. In: *Physical Review E* 70.3 (2004), p. 036120.

[113] Frank Kwasniok. „Analysis and modelling of glacial climate transitions using simple dynamical systems“. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1991 (2013), p. 20110472.

[114] Steven J Lade. „Finite sampling interval effects in Kramers–Moyal analysis“. In: *Physics Letters A* 373.41 (2009), pp. 3705–3709.

[115] Salim Lahmiri. „Multi-scaling analysis of the S&P500 under different regimes in wavelet domain“. In: *International Journal of Strategic Decision Sciences (IJSDS)* 5.2 (2014), pp. 43–55.

[116] Paul Langevin. „Sur la théorie du mouvement brownien“. In: *Compt. Rendus* 146 (1908), pp. 530–533.

[117] Klaus Lehnertz, Lina Zabawa, and M Reza Rahimi Tabar. „Characterizing abrupt transitions in stochastic dynamics“. In: *New Journal of Physics* 20.11 (2018), p. 113043.

[118] Jake Lever, Martin Krzywinski, and Naomi Altman. „Points of significance: Principal component analysis“. In: *Nature methods* 14.7 (2017), pp. 641–643.

[119] Andy K Li, Houk Paek, and Jin-Yi Yu. „The changing influences of the AMO and PDO on the decadal variation of the Santa Ana winds“. In: *Environmental Research Letters* 11.6 (2016), p. 064019.

[120] Decai Li, Min Han, and Jun Wang. „Chaotic time series prediction based on a novel robust echo state network“. In: *IEEE Transactions on Neural Networks and Learning Systems* 23.5 (2012), pp. 787–799.

[121] Shuanglin Li, Judith Perlwitz, Xiaowei Quan, and Martin P Hoerling. „Modelling the influence of North Atlantic multidecadal warmth on the Indian summer rainfall“. In: *Geophysical Research Letters* 35.5 (2008).

[122] Shujun Li, Lixin Wu, Yun Yang, Tao Geng, Wenju Cai, Bolan Gan, Zhaohui Chen, Zhao Jing, Guojian Wang, and Xiaohui Ma. „The Pacific Decadal Oscillation less predictable under greenhouse warming“. In: *Nature Climate Change* 10.1 (2020), pp. 30–34.

[123] Zhen Li, Xin Bian, Xiantao Li, and George Em Karniadakis. „Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism“. In: *The Journal of chemical physics* 143.24 (2015), p. 243128.

[124] Zhen Li, Hee Sun Lee, Eric Darve, and George Em Karniadakis. „Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: Application to polymer melts“. In: *The Journal of chemical physics* 146.1 (2017), p. 014104.

[125] Kevin K Lin and Fei Lu. „Data-driven model reduction, Wiener projections, and the Koopman-Mori-Zwanzig formalism“. In: *Journal of Computational Physics* 424 (2021), p. 109864.

[126] Johannes Lohmann and Peter Ditlevsen. „Variability, correlation and forcing of Dansgaard-Oeschger cycles constrain underlying mechanism.“ In: *Geophysical Research Abstracts*. Vol. 21. 2019.

[127] Johannes Lohmann and Peter D Ditlevsen. „Random and externally controlled occurrences of Dansgaard–Oeschger events“. In: *Climate of the Past* 14.5 (2018), pp. 609–617.

[128] Harvard Lomax, Thomas H Pulliam, David W Zingg, Thomas H Pulliam, and David W Zingg. *Fundamentals of computational fluid dynamics*. Vol. 246. Springer, 2001.

[129] Erick López, Carlos Valle, Héctor Allende, Esteban Gil, and Henrik Madsen. „Wind power forecasting based on echo state networks and long short-term memory“. In: *Energies* 11.3 (2018), p. 526.

[130] Edward N Lorenz. „Deterministic nonperiodic flow“. In: *Journal of atmospheric sciences* 20.2 (1963), pp. 130–141.

[131] Zhenghui Lu, Naiming Yuan, Qing Yang, Zhuguo Ma, and Jürgen Kurths. „Early warning of the Pacific Decadal Oscillation phase transition using complex network analysis“. In: *Geophysical Research Letters* 48.7 (2021), e2020GL091674.

[132] Josef Ludescher, Avi Gozolchiani, Mikhail I Bogachev, Armin Bunde, Shlomo Havlin, and Hans Joachim Schellnhuber. „Improved El Niño forecasting by cooperativity detection“. In: *Proceedings of the National Academy of Sciences* 110.29 (2013), pp. 11742–11745.

[133] Mantas Lukoševičius. „A practical guide to applying echo state networks“. In: *Neural networks: Tricks of the trade*. Springer, 2012, pp. 659–686.

[134] Mantas Lukoševičius and Herbert Jaeger. „Reservoir computing approaches to recurrent neural network training“. In: *Computer Science Review* 3.3 (2009), pp. 127–149.

[135] Mantas Lukoševičius, Herbert Jaeger, and Benjamin Schrauwen. „Reservoir computing trends“. In: *KI-Künstliche Intelligenz* 26.4 (2012), pp. 365–371.

[136] Jing-Jia Luo, Sébastien Masson, Swadhin K Behera, and Toshio Yamagata. „Extended ENSO predictions using a fully coupled ocean–atmosphere model“. In: *Journal of Climate* 21.1 (2008), pp. 84–93.

[137] Andrew J Majda, Ilya Timofeyev, and Eric Vanden Eijnden. „Models for stochastic climate prediction“. In: *Proceedings of the National Academy of Sciences* 96.26 (1999), pp. 14687–14691.

[138] Andrew J Majda, Ilya Timofeyev, and Eric Vanden Eijnden. „A mathematical framework for stochastic climate models“. In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 54.8 (2001), pp. 891–974.

[139] Nishant Malik, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths. „Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks“. In: *Climate Dynamics* 39.3-4 (2012), pp. 971–987.

[140] Nathan J Mantua and Steven R Hare. „The Pacific decadal oscillation“. In: *Journal of oceanography* 58.1 (2002), pp. 35–44.

[141] Nathan J Mantua, Steven R Hare, Yuan Zhang, John M Wallace, and Robert C Francis. „A Pacific interdecadal climate oscillation with impacts on salmon production“. In: *Bulletin of the american Meteorological Society* 78.6 (1997), pp. 1069–1080.

[142] Xuerong Mao. „The truncated Euler–Maruyama method for stochastic differential equations“. In: *Journal of Computational and Applied Mathematics* 290 (2015), pp. 370–384.

[143] Nigel D Marsh and Peter D Ditlevsen. „Climate during glaciation and deglaciation identified through chemical tracers in ice-cores“. In: *Geophysical research letters* 24.11 (1997), pp. 1319–1322.

[144] Cristian Martinez-Villalobos, Matthew Newman, Daniel J Vimont, Cécile Penland, and J David Neelin. „Observed El Niño-La Niña Asymmetry in a Linear Model“. In: *Geophysical Research Letters* 46.16 (2019), pp. 9909–9919.

[145] Frank J Massey Jr. „The Kolmogorov-Smirnov test for goodness of fit“. In: *Journal of the American statistical Association* 46.253 (1951), pp. 68–78.

[146] Larry Medsker and Lakhmi C Jain. *Recurrent neural networks: design and applications*. CRC press, 1999.

[147] Gerald A Meehl, Haiyan Teng, and Julie M Arblaster. „Climate model simulations of the observed early-2000s hiatus of global warming“. In: *Nature Climate Change* 4.10 (2014), pp. 898–902.

[148] Boudjelal Meftah, Olivier Lézoray, and Abdelkader Benyettou. „Novel approach using echo state networks for microscopic cellular image segmentation“. In: *Cognitive Computation* 8.2 (2016), pp. 237–245.

[149] Jun Meng, Jingfang Fan, Yosef Ashkenazy, Armin Bunde, and Shlomo Havlin. „Forecasting the magnitude and onset of El Niño based on climate network“. In: *New Journal of Physics* 20.4 (2018), p. 043036.

[150] William J Merryfield, Woo-Sung Lee, George J Boer, Viatcheslav V Kharin, John F Scinocca, Gregory M Flato, RS Ajayamohan, John C Fyfe, Youmin Tang, and Saroja Polavarapu. „The Canadian seasonal to interannual prediction system. Part I: Models and initialization“. In: *Monthly weather review* 141.8 (2013), pp. 2910–2945.

[151] George S Misyris, Andreas Venzke, and Spyros Chatzivasileiadis. „Physics-informed neural networks for power systems“. In: *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE. 2020, pp. 1–5.

[152] Melanie Mitchell and Mark Newman. „Complex systems theory and evolution“. In: *Encyclopedia of evolution* 1 (2002), pp. 1–5.

[153] W Moon and JS Wettlaufer. „On the interpretation of Stratonovich calculus“. In: *New Journal of Physics* 16.5 (2014), p. 055017.

[154] Hazime Mori. „Transport, collective motion, and Brownian motion“. In: *Progress of theoretical physics* 33.3 (1965), pp. 423–455.

[155] Elizbar A Nadaraya. „On estimating regression“. In: *Theory of Probability & Its Applications* 9.1 (1964), pp. 141–142.

[156] Matthew Newman. „Interannual to decadal predictability of tropical and North Pacific sea surface temperatures“. In: *Journal of climate* 20.11 (2007), pp. 2333–2356.

[157] Matthew Newman, Michael A Alexander, Toby R Ault, Kim M Cobb, Clara Deser, Emanuele Di Lorenzo, Nathan J Mantua, Arthur J Miller, Shoshiro Minobe, Hisashi Nakamura, et al. „The Pacific decadal oscillation, revisited“. In: *Journal of Climate* 29.12 (2016), pp. 4399–4427.

[158] Markus Niemann, Thomas Laubrich, Eckehard Olbrich, and Holger Kantz. „Usage of the Mori-Zwanzig method in time series analysis“. In: *Physical Review E* 77.1 (2008), p. 011117.

[159] Mohammad Norouzi, David J Fleet, and Russ R Salakhutdinov. „Hamming distance metric learning“. In: *Advances in neural information processing systems* 25 (2012).

[160] Keiron O'Shea and Ryan Nash. „An introduction to convolutional neural networks“. In: *arXiv preprint arXiv:1511.08458* (2015).

[161] Yuko M Okumura and Clara Deser. „Asymmetry in the duration of El Niño and La Niña“. In: *Journal of Climate* 23.21 (2010), pp. 5826–5843.

[162] Odd Helge Otterå, Mats Bentsen, Helge Drange, and Lingling Suo. „External forcing as a metronome for Atlantic multidecadal variability“. In: *Nature Geoscience* 3.10 (2010), pp. 688–694.

[163] Milan Paluš, David Hartman, Jaroslav Hlinka, and Martin Vejmelka. „Discerning connectivity from dynamics in climate networks“. In: *Nonlinear Processes in Geophysics* 18.5 (2011), pp. 751–763.

[164] Hae-Jeong Park and Karl Friston. „Structural and Functional Brain Networks: From Connections to Cognition“. In: *Science* 342.6158 (2013), p. 1238411.

[165] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. „On the difficulty of training recurrent neural networks“. In: *International conference on machine learning*. PMLR. 2013, pp. 1310–1318.

[166] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. „Epidemic processes in complex networks“. In: *Reviews of modern physics* 87.3 (2015), p. 925.

[167] Jaideep Pathak, Zhixin Lu, Brian R Hunt, Michelle Girvan, and Edward Ott. „Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data“. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.12 (2017), p. 121102.

[168] RF Pawula. „Approximation of the linear Boltzmann equation by the Fokker-Planck equation“. In: *Physical review* 162.1 (1967), p. 186.

[169] Cecile Penland. „Random forcing and forecasting using principal oscillation pattern analysis“. In: *Monthly Weather Review* 117.10 (1989), pp. 2165–2185.

[170] Cécile Penland and Theresa Magorian. „Prediction of Niño 3 sea surface temperatures using linear inverse modeling“. In: *Journal of Climate* 6.6 (1993), pp. 1067–1076.

[171] Josep Perelló, JM Porra, Miquel Montero, and Jaume Masoliver. „Black–Scholes option pricing within Itô and Stratonovich conventions“. In: *Physica A: Statistical Mechanics and its Applications* 278.1-2 (2000), pp. 260–274.

[172] Sierra Victoria Petersen, Daniel P Schrag, and Peter U Clark. „A new mechanism for Dansgaard-Oeschger cycles“. In: *Paleoceanography* 28.1 (2013), pp. 24–30.

[173] S George H Philander. „El Nino southern oscillation phenomena“. In: *Nature* 302.5906 (1983), pp. 295–301.

[174] R. Quian Quiroga, Thomas Kreuz, and Peter Grassberger. „Event synchronization: a simple and fast method to measure synchronicity and time delay patterns“. In: *Physical Review E* 66.4 (2002), p. 041904.

[175] Stefan Rahmstorf. „Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle“. In: *Nature* 378.6553 (1995), pp. 145–149.

[176] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. „Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations“. In: *Journal of Computational Physics* 378 (2019), pp. 686–707.

[177] Sune O Rasmussen, Matthias Bigler, Simon P Blockley, Thomas Blunier, Susanne L Buchardt, Henrik B Clausen, Ivana Cvijanovic, Dorthe Dahl-Jensen, Sigfus J Johnsen, Hubertus Fischer, et al. „A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy“. In: *Quaternary Science Reviews* 106 (2014), pp. 14–28.

[178] Stephan Rasp and Sebastian Lerch. „Neural networks for postprocessing ensemble weather forecasts“. In: *Monthly Weather Review* 146.11 (2018), pp. 3885–3900.

[179] Stephan Rasp, Michael S Pritchard, and Pierre Gentine. „Deep learning to represent subgrid processes in climate models“. In: *Proceedings of the National Academy of Sciences* 115.39 (2018), pp. 9684–9689.

[180] NAA Rayner, De E Parker, EB Horton, Chris K Folland, Lisa V Alexander, DP Rowell, EC Kent, and A Kaplan. „Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century“. In: *Journal of Geophysical Research: Atmospheres* 108.D14 (2003).

[181] Philippe Renard, Andres Alcolea, and David Ginsbourger. „Stochastic versus deterministic approaches“. In: *Environmental modelling: Finding simplicity in complexity* (2013), pp. 133–149.

[182] Jo Roberts and Thomas D Roberts. „Use of the Butterworth low-pass filter for oceanographic data“. In: *Journal of Geophysical Research: Oceans* 83.C11 (1978), pp. 5510–5514.

[183] W Rüemelin. „Numerical treatment of stochastic differential equations“. In: *SIAM Journal on Numerical Analysis* 19.3 (1982), pp. 604–613.

[184] Jakob Runge, Vladimir Petoukhov, Jonathan F Donges, Jaroslav Hlinka, Nikola Jajcay, Martin Vejmelka, David Hartman, Norbert Marwan, Milan Paluš, and Jürgen Kurths. „Identifying causal gateways and mediators in complex spatio-temporal systems“. In: *Nature communications* 6.1 (2015), pp. 1–10.

[185] Jakob Runge et al. „Inferring causation from time series in Earth system sciences“. In: *Nature Communications* 10.1 (2019), p. 2553.

[186] Francisco Sahli Costabal, Yibo Yang, Paris Perdikaris, Daniel E Hurtado, and Ellen Kuhl. „Physics-informed neural networks for cardiac activation mapping“. In: *Frontiers in Physics* 8 (2020), p. 42.

[187] Maxi San Miguel, Jeffrey H Johnson, Janos Kertesz, Kimmo Kaski, Albert Dıaz-Guilera, Robert S MacKay, Vittorio Loreto, Péter Érdi, and Dirk Helbing. „Challenges in complex systems science“. In: *The European Physical Journal Special Topics* 214.1 (2012), pp. 245–271.

[188] Niklas Schneider and Bruce D Cornuelle. „The forcing of the Pacific decadal oscillation“. In: *Journal of Climate* 18.21 (2005), pp. 4355–4373.

[189] Viktor Sebestyén, Tımea Czvetkó, and János Abonyi. „The applicability of Big Data in climate change research: the importance of system of systems thinking“. In: *Frontiers in Environmental Science* 9 (2021), p. 70.

[190] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. „Activation functions in neural networks“. In: *towards data science* 6.12 (2017), pp. 310–316.

[191] Steven Shreve. *Stochastic calculus for finance I: the binomial asset pricing model*. Springer Science & Business Media, 2005.

[192] Alexander F Siegenfeld and Yaneer Bar-Yam. „An introduction to complex systems science and its applications“. In: *Complexity* 2020 (2020).

[193] Jonatan F Siegmund, Tanja GM Sanders, Ingo Heinrich, Ernst van der Maaten, Sonia Simard, Gerhard Helle, and Reik V Donner. „Meteorological drivers of extremes in daily stem radius variations of beech, oak, and pine in northeastern Germany: an event coincidence analysis“. In: *Frontiers in Plant Science* 7 (2016), p. 733.

[194] Amanpreet Singh, Narina Thakur, and Aakanksha Sharma. „A review of supervised machine learning algorithms“. In: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. Ieee. 2016, pp. 1310–1315.

[195] Abhishekh Srivastava and Timothy DelSole. „Decadal predictability without ocean dynamics“. In: *Proceedings of the National Academy of Sciences* 114.9 (2017), pp. 2177–2182.

[196] Clara Stegehuis, Remco Van Der Hofstad, and Johan SH Van Leeuwaarden. „Epidemic spreading on complex networks with community structures“. In: *Scientific reports* 6.1 (2016), pp. 1–7.

[197] Jochen J Steil. „Backpropagation-decorrelation: online recurrent learning with O (N) complexity“. In: *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*. Vol. 2. IEEE. 2004, pp. 843–848.

[198] Ralf Steuer, Jürgen Kurths, Carsten O Daub, Janko Weise, and Joachim Selbig. „The mutual information: detecting and evaluating dependencies between variables“. In: *Bioinformatics* 18.suppl_2 (2002), S231–S240.

[199] Rouslan L Stratonovich. *Topics in the theory of random noise*. Vol. 2. CRC Press, 1967.

[200] Courtenay Strong and Gudrun Magnusdottir. „The role of tropospheric Rossby wave breaking in the Pacific decadal oscillation“. In: *Journal of Climate* 22.7 (2009), pp. 1819–1833.

[201] Emma B Suckling, Geert Jan van Oldenborgh, Jonathan M Eden, and Ed Hawkins. „An empirical model for probabilistic decadal prediction: global attribution and regional hindcasts“. In: *Climate Dynamics* 48.9-10 (2017), pp. 3115–3138.

[202] Philip Sura, Matthew Newman, Cécile Penland, and Prashant Sardeshmukh. „Multiplicative noise and non-Gaussianity: A paradigm for atmospheric regimes?“ In: *Journal of the atmospheric sciences* 62.5 (2005), pp. 1391–1409.

[203] Rahimi Tabar. *Analysis and data-based reconstruction of complex nonlinear dynamical systems*. Vol. 730. Springer, 2019.

[204] Stefan Thurner, Rudolf Hanel, and Peter Klimek. *Introduction to the theory of complex systems*. Oxford University Press, 2018.

[205] Robert Tibshirani. „Regression shrinkage and selection via the lasso“. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288.

[206] Axel Timmermann, Hezi Gildor, Michael Schulz, and Eli Tziperman. „Coherent resonant millennial-scale climate oscillations triggered by massive meltwater pulses“. In: *Journal of Climate* 16.15 (2003), pp. 2569–2585.

[207] Matthew H Tong, Adam D Bickett, Eric M Christiansen, and Garrison W Cottrell. „Learning grammatical structure with echo state networks“. In: *Neural networks* 20.3 (2007), pp. 424–432.

[208] Anastasios A Tsonis and Paul J Roebber. „The architecture of the climate network“. In: *Physica A: Statistical Mechanics and its Applications* 333 (2004), pp. 497–504.

[209] Robert Vautard, Pascal Yiou, and Michael Ghil. „Singular-spectrum analysis: A toolkit for short, noisy chaotic signals“. In: *Physica D: Nonlinear Phenomena* 58.1-4 (1992), pp. 95–126.

[210] JM Wallace, EM Rasmusson, TP Mitchell, VE Kousky, ES Sarachik, and H Von Storch. „On the structure and evolution of ENSO-related climate variability in the tropical Pacific: Lessons from TOGA“. In: *Journal of Geophysical Research: Oceans* 103.C7 (1998), pp. 14241–14259.

[211] Shanshan Wang, Jianping Huang, Yongli He, and Yuping Guan. „Combined effects of the Pacific decadal oscillation and El Nino-southern oscillation on global land dry–wet changes“. In: *Scientific reports* 4.1 (2014), pp. 1–8.

[212] Sun-Chong Wang. „Artificial neural network“. In: *Interdisciplinary computing in java programming*. Springer, 2003, pp. 81–100.

[213] Geoffrey S Watson. „Smooth regression analysis“. In: *Sankhyā: The Indian Journal of Statistics, Series A* (1964), pp. 359–372.

[214] Duncan J Watts and Steven H Strogatz. „Collective dynamics of ‘small-world’ networks“. In: *Nature* 393.6684 (1998), p. 440.

[215] Peter J Webster and Song Yang. „Monsoon and ENSO: Selectively interactive systems“. In: *Quarterly Journal of the Royal Meteorological Society* 118.507 (1992), pp. 877–926.

[216] Paul J Werbos. „Backpropagation through time: what it does and how to do it“. In: *Proceedings of the IEEE* 78.10 (1990), pp. 1550–1560.

[217] Frederik Wolf, Jurek Bauer, Niklas Boers, and Reik V Donner. „Event synchrony measures for functional climate network analysis: A case study on South American rainfall dynamics“. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.3 (2020), p. 033102.

[218] L Wu, Z Liu, R Gallimore, R Jacob, D Lee, and Y Zhong. „Pacific decadal variability: The tropical Pacific mode and the North Pacific mode“. In: *Journal of Climate* 16.8 (2003), pp. 1101–1120.

[219] Jining Yan, Lin Mu, Lizhe Wang, Rajiv Ranjan, and Albert Y Zomaya. „temporal convolutional networks for the Advance prediction of enSo“. In: *Scientific Reports* 10.1 (2020), pp. 1–15.

[220] Xiaosong Yang, Anthony Rosati, Shaoqing Zhang, Thomas L Delworth, Rich G Gudgel, Rong Zhang, Gabriel Vecchi, Whit Anderson, You-Soon Chang, Timothy DelSole, et al. „A predictable AMO-like pattern in the GFDL fully coupled ensemble initialization and decadal forecasting system“. In: *Journal of Climate* 26.2 (2013), pp. 650–661.

[221] S-W Yeh and Ben P Kirtman. „The impact of internal atmospheric variability on the North Pacific SST variability“. In: *Climate dynamics* 22.6-7 (2004), pp. 721–732.

[222] Izzet B Yildiz, Herbert Jaeger, and Stefan J Kiebel. „Re-visiting the echo state property“. In: *Neural networks* 35 (2012), pp. 1–9.

[223] Yayun Zheng, Fang Yang, Jinqiao Duan, Xu Sun, Ling Fu, and Jürgen Kurths. „The maximum likelihood climate change for global warming under the influence of greenhouse effect and Lévy noise“. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.1 (2020), p. 013132.

[224] Yuanran Zhu and Daniele Venturi. „Generalized Langevin equations for systems with local interactions“. In: *Journal of Statistical Physics* 178.5 (2020), pp. 1217–1247.

[225] Robert Zwanzig. „Memory effects in irreversible thermodynamics“. In: *Physical Review* 124.4 (1961), p. 983.

[226] Robert Zwanzig. „Nonequilibrium Statistical Mechanics Oxford University Press“. In: *New York* (2001).