On Nonlinear Time Series Analysis and Climate Variability

Krishna Chebolu

November 2023

Abstract

Nonlinear Time Series Analysis is a versatile field that has yet to realize its complete potential. NTSA has been extensively used in Finance and Medicine, but not as much in the natural sciences, such as climate sciences. This paper attempts to answer the question "What is nonlinear time series analysis and how is it used in modeling climate variability and other such phenomena?" We begin with the foundations of NTSA and introduce some common techniques. Then, we highlight prominent applications of NTSA in current climate modeling such as the El Niño Southern Oscillation and Ozone layer depletion. We also discuss the predictive capabilities and shortcomings of various models. Finally, this paper offers the reader an introduction to seminal works, ongoing research, and problems to further explore. A reader can expect to gain familiarity with important concepts in the field and understand their potential applications in problems around the world.

Contents

1 Introduction

1.1 Background and History of Nonlinear Time Series Analysis

Nonlinear Time Series Analysis (NTSA) has emerged as a transformative approach to studying complex dynamical systems, offering insights into diverse fields ranging from physics to economics. The history of NTSA traces back to the work of the French mathematician Henri Poincaré, who, in the late 1800s, explored the dynamics of celestial bodies within the context of his three-body problem. The Poincaré three-body problem refers to the challenge of accurately predicting the future positions and motions of three celestial bodies, such as planets or stars, which exert gravitational forces on each other [Bar97]. Poincaré's discoveries laid the foundation for chaos theory, a pivotal component of NTSA. However, it was not until the latter half of the 20th century that NTSA truly gained momentum, driven by advances in computational technology and a growing recognition of its value across disciplines.

One of the milestones in the development of nonlinear time series analysis was the work of Edward Lorenz in the 1960s. Lorenz's formulation of the famous Lorenz equations revolutionized meteorology and highlighted the sensitivity to initial conditions, a hallmark of chaotic systems [Lor63]. This work was instrumental in guiding researchers towards nonlinear dynamics. Concurrently, the study of fractals by Benoît Mandelbrot in the 1970s further contributed to the understanding of self-similarity and scaling in natural phenomena [MA79].

NTSA gained prominence in climate science during the late 20th century as researchers sought to study the inherently nonlinear behavior of our planet's climate system. The wide-range use of powerful computers enabled the development of sophisticated models for climate prediction, incorporating much-needed nonlinear elements. Climate phenomena like El Niño Southern Oscillation (ENSO), which eluded traditional linear models, became a hotspot for NTSA research. We discuss the ENSO phenomenon in section 4.1. As climate change emerged as a global concern, the demand for accurate, long-term predictions intensified, driving further advancements.

1.2 Climate Variability

Climate variability is a fundamental aspect of Earth's climate system, encompassing natural fluctuations that occur over various timescales, from seasons to millennia. Understanding and characterizing climate variability are essential for predicting weather patterns, assessing long-term climate trends, and addressing environmental challenges. Climate systems exhibit intricate behaviors that extend beyond superficial linear relationships. Nonlinear time series analysis provides a powerful toolkit for investigating these complex interactions and nonlinearities within climate time series data.

Researchers have employed NTSA techniques, including Rescaled Range Analysis (R/S) and Detrended Fluctuation Analysis (DFA), to analyze climate time series data from diverse sources, including temperature records, sea-level measurements, and ice-core data. These methods allow researchers to uncover long-range correlations, persistence, and self-similarity within climate data, offering valuable insights into the Earth's changing climate.

1.3 Scope and Objective

Understanding climate variability is one of the most pressing challenges of our time. Our planet's climate and weather system comprises complex variables and underlying dynamics. Examples of such dynamics include everything from long-term temperature shifts to weather anomalies. These seemingly random changes and fluctuations significantly affect our ecosystems, agriculture, water resources, and, in general, human societies everywhere.

For a long time, we have used linear models, regression, and time series analysis to understand and predict climate variability [17; Cha05]. However, as more variables are being discovered and measured, more relationships in the system surface, making it challenging to study the underlying dynamics with only linear tools, an example of which is illustrated in [EP04]. As research questions in nonlinear dynamics develop relating to climate change, it can be challenging to navigate through the vast number of topics and academic publications.

This paper aims to synthesize the underlying mathematical and physical principles that govern climate variability by examining the current body of research on nonlinear time series analysis. Thus, this paper hopes to be a primer for any reader new to the field.

Section 2 introduces foundational concepts that nonlinear time series analysis relies on, such as state space reconstruction and different mathematical characterization methods, such as Lyapunov exponents and attractors. Section 3 will discuss nonlinear techniques that help conduct analyses that are more robust. Section 4 discusses different analysis methods used to support climate variability studies. This section helps put in perspective some of the abstract concepts discussed in section 2. Some examples used in section 4 are models of the EN$\tilde{\text{S}}$O phenomenon and the ozone layer. Section 5 discusses the predictive capabilities and limitations of nonlinear time series analysis. Section 6 suggests reading material for further exploration and poses some questions that still need examining. We conclude and offer some final remarks in section 7.

1.4 Methodology

This literature review assesses the state of knowledge regarding nonlinear time series analysis and its application to climate variability. This review's research question is "What is the current understanding of NTSA and its relevance in modeling and predicting climate variability or phenomena?" We implement a comprehensive search strategy to address this question, utilizing academic databases such as Google Scholar.

A systematic two-stage screening process selected relevant literature, assessing titles and abstracts for initial relevance and subsequent examination of full-text articles. Data extraction from the selected articles focused on key elements such as authorship, publication year, research methodology, findings, and implications. We analyzed the gathered literature to identify recurring patterns, research trends, and gaps in applying NTSA to climate variability. Quality assessments considering factors such as study designs, methods, and references weighed the reliability and relevance of the selected studies.

Some academic papers from the late 1900s cited in thousands of papers are *base papers*. Base papers were early in the development of a field of study, the go-to papers in a field. Due to their extensive contributions, we will discuss some of the base papers used in this review separately for any reader to explore more. On the contrary, we selected some papers cited less than a hundred times. These papers are niche and refer to many base papers to strengthen their argument or method, making them worth reading and discussing.

2 Foundations of Nonlinear Time Series Analysis

2.1 State Space Reconstruction

A state space is an abstract representation of a dynamic system's complete condition. It is a multidimensional space where each dimension corresponds to a variable that somehow describes the system; for example, temperature measures a part of the weather. A point in the multidimensional space can tell us everything we want to know about the system at that time. If we look at the weather app for a particular time, we can see the temperature, humidity, chance of rain, and so on; the particular time would be the point in space we are looking for. The variables, such as temperature, humidity, and so on, would be the dimensions of our system.

Static systems have unchanging, constant properties, while dynamic systems involve variables and changes over time, making them inherently time-dependent and subject to evolution. We obtain multiple data points for our discussion about NTSA as dynamic systems evolve with time. Temperature would be a series of data points concerning the weather, a dynamic system. These series of data points are a *time series*, analogous to a video that is a continuous series of pictures. When we attempt to plot these data points, they trace a path called a trajectory. These trajectories reveal how a system changes. In dynamic systems like the weather, we cannot perfectly measure all the internal variables because we might not know how many are present. However, reconstructing the dynamic systems from time series can allow us to study their structure more closely [Pac+80]. While the reconstruction may not be identical to the actual dynamics of a system, they will have a similar structure [BK15]. This similarity is advantageous since we can confidently project our findings from the reconstructed system onto the original system.

2.1.1 Delay-Coordinate Embedding

Reconstructing a state space from a single time series uses delay coordinate embedding. We can reconstruct a multidimensional dynamical system from a scalar time series.

Consider a scalar measurement x, say temperature. We can construct an m-dimensional vector $\overrightarrow{R}(t)$ from m time-delayed measurements $x(t)$, such that

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

where t is the time of measurement and τ is the chosen time delay. The vector $\overrightarrow{R}(t)$ represents a series of measurements of x, temperature, in our case, over a certain time period. The time-delay variable τ represents the intervals of these measurements.

Choosing m and τ requires serious thought. For instance, if τ has a small value, the correlation in the m-coordinate data may be incredibly strong. In the opposite case of τ being large, there might not be a significant correlation worth studying. Continuing our temperature example, say that we want to analyze the temperature over the last week to deduce a pattern. For this, we take temperature measurements at particular intervals (τ). We do not have enough data to form a correlation if we take one temperature measurement daily for seven days. If we take a data point every second, there are too many patterns to analyze and make practical, generalized deductions. We choose an in-between case: perhaps a measurement every thirty minutes or hour. This will result in a more manageable and generalized time series.

In the vector space $\overrightarrow{R}(t)$, two implicit assumptions can cause problems in practice. First, the time series is evenly spaced. One way to even out irregular time series is through interpolation, but then we cannot study the actual dynamics of the system. Second, the measurement function produces a smooth function of the original system. We expect reconstruction not to be a replica; thus, there will be rough edges in the function that would have been averaged out with more variables to fill in our knowledge gaps. To mitigate these rough edges, repeating the analysis with different measurements and confirming consistent results provides a smooth function and confidence in results.

2.1.2 Parameter Estimation

While we theoretically know what we need for the values τ, the time delay, and m, the embedding dimension, it can be difficult to obtain them in practice. While analyzing data, we do not always know the dimensions of the entire system. Thus, estimating these parameters is a significant challenge in delay-coordinate embedding. Selecting the time delay variable τ is tricky because there is no universally optimal value, and it often requires trial and error. Overestimating or underestimating τ can lead to information loss or sensitivity to noise, and noise in the data further complicates the selection process. Researchers may use heuristics and data-driven approaches, but nonlinear time series analysis remains a challenging and data-dependent task.

Generally, the τ value should be reasonably small to avoid other problems, such as a reconstructed attractor's (see section 2.2.1 for attractors) trajectories overlap or *folding in on themselves*, making it difficult to study. There are some strategies to calculate τ, but none are universal. One strategy outlined by Bradley and Kantz [BK15] is computing a statistic that measures the independence of τ-separated points in the time series.

After choosing τ, the next step is determining the embedding dimension. The False Near Neighbor (FNN) algorithm proposed by Kennel et al. gives an approach to finding an optimal dimension [KBA92a].

We have data points but are unsure how many dimensions we need to represent them accurately, so we start with a small guess. Say we start with a two-dimensional (2D) space. In this space, we compute the nearest neighbor for each data point. Then, add a dimension, a 3D space. If the relationship between data points changes, as indicated in changes to the nearest neighbor calculations, then the 2D space was not enough to represent accurately the data. Then, we add another dimension and repeat the process in the 4D space. We do this until relationship differences stabilize or go lower than a threshold. The FNN algorithm suggests adding extra dimensions [RM97].

2.2 Mathematical Characterization

2.2.1 Attractors

Attractors are mathematical constructs that provide insight into the long-term evolution of dynamic systems. They unveil recurring patterns and trajectories that govern their behavior over time. An attractor is a subset of a system's state space to which a system's trajectories converge or stabilize. There are three kinds of attractors:

1. Fixed-point attractors are equilibrium states where a system converges to a single stable point, representing rest.

2. Periodic attractors are stable but not at rest. These attractors represent repeating patterns or cycles, such as periodic orbits.

3. Strange attractors are complex, non-repeating attractors found in chaotic systems. Their behavior can appear random but is deterministic.

Strange attractors are a hallmark of chaos and exhibit extreme sensitivity to initial conditions. Turbulence is a prime example of such behavior. When we observe turbulent motion, there is motion at multiple frequencies, meaning that many kinds of motion, fast and slow, occur simultaneously. In turbulence, a nonperiodic or periodic motion with an infinite period was observed [TE89]. It is akin to attempting to predict the path of a leaf caught in the wind; while there may be some semblance of repetition, the prediction periods are practically beyond our consideration.

It became evident that a simple dynamical system could not produce such a motion. Lorenz first notices a system of this kind [Lor63]. The system gives an approximate description of a fluid layer heated from below. The Lorenz attractor utilizes the equations below

$$\begin{aligned} dx/dt &= -ax + ay \\ dy/dt &= -xz + bx - y \\ dz/dt &= -xy - cz \end{aligned}$$

where x is proportional to the intensity of the convective motion, y is proportional to the horizontal temperature variation, z is proportional to the vertical temperature variation, and a, b, and c are constants. Since the curve is deterministic, we can plot these equations.

Like most concepts in this paper, we use attractors to study the weather and climate. In their paper about attractors and the weather, Tsonis and Elsner note that while the atmosphere is chaotic and difficult to reconstruct, its evolution is limited and confined to a specific area that the attractor occupies [TE89]. This limits our field of study, saving time and computational power.

2.2.2 Lyapunov Exponents

Lyapunov exponents, named after the Russian mathematician Aleksandr Lyapunov, offer valuable insights into the behavior of dynamic systems. These values quantify the sensitivity of a dynamical system to its initial conditions. A positive largest Lyapunov exponent, $\lambda_1 > 0$, signifies chaotic behavior within the system. In such cases, a minor alteration in the initial conditions can lead to a substantial divergence in the trajectories. Conversely, a largest negative Lyapunov exponent, $\lambda_1 < 0$, suggests stability. Finally, when the largest Lyapunov exponent is 0, $\lambda_1 = 0$, the system is considered marginally stable or neutral; changes in initial conditions have limited long-term influence.

There are limitations to relying solely on Lyapunov exponents. They assume that a system is continuous, deterministic, and that we know the underlying equations. Deterministic means a system's future state or outcome is entirely predictable based on its initial conditions and known rules or laws. Real-world systems may deviate from these assumptions frequently. More discussion about Lyapunov exponents is in Dingwell's work [Din06].

There are many algorithms to compute Lyapunov exponents. One widely used algorithm to calculate the exponents is the Rosenstein algorithm [RCD93]. This algorithm operates in the following steps:

1. Embedding: To analyze the system's dynamics, first, we reconstruct the state space using delay-coordinate embedding.

2. Nearest neighbor search: For each point in this new data set, we calculate the distance to all other points and find the nearest neighbor. For each state X_i, the closest neighbor X_j in phase space greater than the time series' mean period μ, is eligible, X_e. The distance is calculated as $d_e = ||X_i - X_i||$, where $i - e > \mu$.

3. Logarithmic growth: As we iterate through the time series data, we track how the distance between data points evolves. In a chaotic system, nearby points tend to diverge over time, and we measure this divergence as a function of time.

4. Estimation: We estimate the largest Lyapunov exponent λ by analyzing the rate of this divergence. It quantifies how quickly nearby trajectories in the phase space move apart. In a chaotic system, this rate is positive, indicating sensitive dependence on initial conditions.

5. Finding the exponent: We analyze the growth of the distances over time and take the slope of a graph showing the natural logarithm of these distances. The slope represents the largest Lyapunov exponent.

Wolf et al. elaborate on various methods to calculate Lyapunov exponents, see [Wol+85].

2.2.3 Fractals

Fractals are abstract geometric shapes or structures that exhibit self-similarity. If you magnify a fractal to observe a part of the structure, it will look identical to the overarching structure, which happens no matter how much one zooms in; this property is *self-similarity*. Fractals have a way of appearing naturally all around us [Bar+88]. In nonlinear time series analysis, fractals and their analysis play an important role in characterizing self-similarity found in turbulent systems such as the climate [Sre91].

Fractal analysis helps us identify whether observed behaviors correlate with the rest of the structure. Thus, identifying discrepancies becomes a slightly more manageable task. Fractal dimensions provide insight into these structures. Unlike traditional geometry, fractal dimensions are not confined to integer values. Instead, they quantify the complexity of a pattern or structure. For instance, the renowned Koch snowflake boasts a fractal dimension of 1.2619.

The Hurst exponent, H, represents the 'memory' or 'persistence' of time series data. Imagine a series of numbers that goes up and down; the Hurst exponent tells us whether these ups and downs are just random jumps or if there's a hidden pattern that keeps pushing things in the same direction. We can link the Hurst exponent to fractals through self-similarity and long-range dependence.

Fractals are complex shapes that display self-similarity, with similar patterns repeating at various scales. Similarly, the Hurst exponent quantifies how time series data exhibits similar patterns at different time scales. Moreover, fractals and the Hurst exponent are associated with long-range dependence, where events or data points at one location can significantly affect those at distant points. This connection emphasizes the role of the Hurst exponent in characterizing self-similarity and long-range dependence in data, much like how fractals exhibit these properties in geometric shapes.

A high Hurst exponent suggests strong long-term dependencies or trends in the data, while a low exponent implies more randomness. This concept is particularly valuable when examining time series data in fields like climate science or finance, where understanding patterns and trends over time is crucial. We illustrate this concept in a more detailed manner in section 4.2.1.

3 Nonlinear Techniques

3.1 Surrogate Data

Before nonlinear time series analysis is used, we must determine whether a dynamic system is chaotic. Detecting chaotic behavior of a system within time series data can be challenging, mainly when dealing with short and noisy time series data. These circumstances can confound dimension estimates and Lyapunov exponent measurements, potentially leading to false conclusions regarding the presence of chaotic behavior.

To address this challenge, statistical hypothesis testing is a valuable tool. It allows us to assess whether the observed time-series data is more likely to represent nonlinear, chaotic dynamics or simply random noise. This is crucial, as finite-length time series can originate either from a noise process or from a low-dimensional deterministic process [The+92].

Focusing on detecting nonlinear structures within time series data is more practical than characterizing them comprehensively. Estimating the dimension of the system and interpreting its value is computationally demanding and prone to errors, primarily because we might not have complete knowledge of the underlying dynamics.

Statistical hypothesis testing involves formulating a null hypothesis, against which we compare our observations using a specific statistic. The goal is to reject the null hypothesis by demonstrating that the statistic quantifies a property of the time series data that is inconsistent with the null hypothesis. In nonlinear time series analysis, test statistics' probability distribution under a simple null hypothesis is usually unknown. We turn to *surrogate data* to perform hypothesis testing effectively.

We generate surrogate data carefully to mimic the conditions of the null hypothesis while retaining certain characteristics of the original time series, such as its statistical properties such as mean and variance. Surrogate data helps mitigate the risk of false positives in chaotic dynamics detection. Without surrogate data, noise or random fluctuations in the data might erroneously lead to the conclusion of chaos. Surrogate data provides a baseline for randomness, making it less likely to misinterpret noise as chaos. We can generate multiple sets of surrogate data to assess the robustness of conclusions. If statistical tests consistently reject the null hypothesis across various surrogate datasets, it strengthens the confidence in the presence of nonlinearity and chaotic dynamics in the original data. See Theiler et al. and Schreiber and Schmitz [The+92; SS00] for an elaborate discussion on algorithms and null hypotheses concerning surrogate data and Lucio et al. [LVR12] for recent developments.

3.2 Recurrence Plots

Recurrence plots (RPs) are data analysis tools that help us understand time series data better through two-dimensional visualization, first introduced by Eckmann et al [EKR+95]. These plots are particularly useful because they allow us to spot repeating patterns and behaviors often occurring in complex systems. Imagine them as magnifying glasses for time series data, helping us zoom in on those recurrent phenomena that might not be immediately obvious. One of their most valuable applications is in studying chaotic systems—those systems that seem unpredictable at first glance. Traditional methods of analyzing time series data may struggle with chaos, but recurrence plots shine here. They help researchers dig deep into the system's dynamics, pinpoint transitions between different states, and reveal the underlying attractors.

We use the following steps to create an RP.

1. Consider a time series. We use delay-coordinate embedding on it to reconstruct it into a multidimensional space.

2. We construct a recurrent matrix $R_{ij}(\epsilon)$, where ϵ is a pre-chosen threshold value to determine how close a pair of points are. The matrix is generated as follows

$$R_{ij}(\epsilon) = \theta(\epsilon - ||\overrightarrow{x}_i - \overrightarrow{x}_j||)$$

Here $\overrightarrow{x}$ represents a randomly chosen data point. The distance between two points is represented by $\overrightarrow{x}_i - \overrightarrow{x}_j$. If the chosen ϵ is greater than the distance, it is close. The θ function assigns 1 if the points are close, and 0 if not.

3. We plot this recurrent matrix. A value of 1 is a black dot and a value of 0 is white.

Recurrence plots look like square matrices, and each cell in the matrix corresponds to a pair of time points from the original time series. If there is a dot or a specific color in a cell, it means that the data points at those time points are recurrent, based on the threshold, ϵ we set. The diagonal line running from the top-left to the bottom-right of the matrix represents self-recurrences—instances when a state or pattern repeats itself. When you see clusters of dots or patterns close together in the plot, it indicates areas of frequent recurrence. These clusters often represent stable behaviors or attractors within the system. Conversely, when you notice gaps or empty regions in the plot, it means there are periods when the system transitions between different states or behaviors.

4 Applications to Climate Variability

4.1 El Niño-Southern Oscillation

El Niño-Southern Oscillation, or ENSO, is a disruption in winds and sea surface temperatures (SST) affecting tropical and subtropical climates in the Pacific Ocean every few years. The British scientist Sir Gilbert Walker discovered this through a difference in barometer readings while on an assignment in India. Sir Walker noticed that a pressure rise in the east resulted in a drop in the west and vice versa. Walker coined the term Southern Oscillation to capture these dramatic fluctuations. ENSO links to many health concerns in South Asia, with a few studies in Bangladesh.

Its oscillatory behavior characterizes ENSO, alternating between two primary phases: *El Niño* and *La Niña*. El Niño is the warming phase of the sea; it weakens trade winds and alters rainfall patterns in various regions worldwide. La Niña is the cooling phase. La Niña intensifies trade wings and can cause the opposite effects in rainfall patterns as El Niño. The modeling of these phases and their transitions has been at the intersection of climate variability and nonlinear time series analysis. This section will review a few methods used to capture this phenomenon.

4.1.1 Smooth Transition Autoregressive Model

Autoregression, a prevalent time series modeling technique, relies on past observations to predict future values. However, this method inherently assumes that the future will closely resemble the past; potentially failing to account for ENSO's abrupt changes or rapid developments. The Smooth Transition Autoregressive (STAR) time series method represents an advanced iteration of traditional autoregressive models. Teräsvirta devised this method that is suitable for series with cyclical variations and turbulent periods, such as ENSO [TA92]. Unlike conventional autoregressive models, STAR models incorporate a mechanism for adapting to evolving conditions, rendering them highly effective for modeling intricate time series data.

A model for ENSO, as described by Hall et al. [HST01], is:

$$y_t = \pi_{10} + \pi_1' w_t + (\pi_{20} + \pi_2' w_t) F(y_{t-d}) + u_t$$

where $\pi_j = (\pi_{j1}, ..., \pi_{jp})'; j = 1, 2; w_t = (y_{t-1}, ..., y_{t-p})'$; and $u_t \sim NID(o, \sigma_u^2)$.

The left-hand side of the equality is what we wish to model, denoted by y_t. When there are no transitions, the intercepts of the function are π_{10} and π_{20}. The term $\pi_1' w_t$ is an autoregressive component. It is the dot product of the lagged values w_t and their weights π_1'. This captures the dependence of y_t on its past values. Similarly, $\pi_2' w_t$ is another autoregressive component.

The transition function is $F(y_{t-d})$; this is the primary component of the mechanism that makes this model more effective for time series. A lagged value denoted by y_{t-d} passes through a logistic or exponential function to enable a smooth transition between regimes. In our equation above, the terms $\pi_1' w_t$ and $\pi_2' w_t$ represent the behavior of the time series in different regimes. The final term u_t represents the random variability or noise in the data.

In the STAR models, the transition function is crucial. The logistic transition function of order p is:

$$F(y_{t-d}) = (1 + e^{-\gamma_L (y_{(t-d)} - c_L)})^{-1}$$

where $\gamma_L > 0$. The model is called logistic STAR or LSTAR when this function is used. The positive parameter γ_L determines the steepness of the transition. A higher value translates to a more abrupt transition, and vice versa. The critical point c_L is when the transition occurs.

In the exponent, if $(y_{t-d} - c_L) = 0$, then the function's output is 0.5, indicating an equal probability of transitioning between states. The function $F(y_{t-d})$ returns a value in the range [0,1].

The exponential transition function of order is given by

$$F(y_{t-d}) = 1 - e^{-\gamma_E (y_{(t-d)} - c_E)^2}$$

where $\gamma_E > 0$. The model is called exponential STAR or ESTAR when this function is used. γ_E is this function's equivalent of γ_L from the LSTAR transition function. The parameter c_E represents the center or location of the transition along y_{t-d}. For more discussion of this model, see [TA92; Ter94] and references within.

Employing the LSTAR model described earlier, Hall et al. achieved a deeper comprehension of turbulent disturbances than the model's linear counterparts did. However, due to its lack of stability, the model is limited in its forecasting capabilities, although it can still provide short-term predictions. The authors acknowledged the inherent complexity of systems like ENSO and suggested that augmenting the model with additional variables could enhance its performance.

4.1.2 Singular Spectrum Analysis

It is almost impossible to obtain the perfect time series. More often than not, we can only study short and noisy time series that we cannot interpret with conventional tools. Thus, we use singular spectrum analysis (SSA) to extract as much reliable information as possible without prior knowledge about the underlying system. This property is unique to SSA and is called *signal-to-noise enhancement*.

SSA begins by representing the time series as a trajectory in a high-dimensional vector space. This representation involves creating overlapping views of the time series data utilizing a *sliding window*, a moving frame that covers a time series segment at each step. We do this by using the embedding and delay-coordinate techniques discussed earlier. After reconstruction, singular value decomposition techniques decompose the trajectory matrix into its almost-fundamental components. These components include trends, oscillatory patterns, and noise. Then, we can attempt to isolate the noise and mitigate a significant amount of its influence on the time series data. We can highlight other components; for instance, we can isolate a particular pattern and reconstruct the data influenced by identified pattern. See [Ghi+02] and references therein for an elaborate discussion on more spectral methods in nonlinear time series and nonlinear dynamics.

Once SSA breaks down a time series into different components, one can compare these components with components of other time series to find a correlation. One study compared time series data from the ENSO phenomenon and cholera frequency in Bangladesh [Pas+00]. Their study found significant similarities in the dominant frequency of interannual variability of cholera cases and ENSO. This led to the conclusion that ENSO, and possibly other climate variability phenomena, influenced how cholera grew in Bangladesh.

Similar applications of SSA are in other climatic time series data as well [VG89; MVG98]. McGregor and Ebi did another elaborate study linking ENSO to various health in various parts of the world in [ME18].

4.2 Ozone Layer

The ozone (O_3) layer is critical to Earth's atmosphere located in the stratosphere. The ozone layer is a protective layer that absorbs and filters out a significant portion of the sun's harmful ultraviolet radiation. Excessive radiation harms aquatic life and can cause adverse effects on humans, such as skin cancer. The discovery of the ozone layer's importance and its depletion led to international efforts to preserve the layer's vitality, such as the Montreal Protocol [Vel+07].

The layer's depletion occurred over time; thus, various methods from time series analysis offered insights into understanding the ozone layer's decline and improvement. We will look at a few of those methods.

4.2.1 Rescaled Range Analysis

Rescaled Range Analysis (R/S), also known as the Hurst exponent analysis, analyzes and characterizes the long-term memory or persistence in time series data. British hydrologist Harold Hurst developed this analysis to study the rate of river flow [Hur51]. This method uses a unique value called the R/S statistic.

We can compute the R/S statistic by dividing the range of the cumulative sum of the data by its standard deviation. To assess the memory in the data after obtaining our time series, we compute the R/S statistic for different window sizes in the data. We plot the various computed R/S statistics against their window sizes on a log-log scale. Finally, a linear regression line on the log-log plot estimates the Hurst exponent, H. If $H > 0.5$, future values will likely follow past observed trends. If $H = 0.5$, then it is random. If $H < 0.5$, future values might behave conversely to the past observed values, referred to as anti-persistence. There can be different thresholds for different studies.

A study by Jan et al. about the ozone layer fluctuation in Pakistan used this analysis method [Jan+14]. The study's authors found anti-persistence, indicating that the ozone layer's future would behave oppositely as it did in the past. This particular study used time series data from 1970 - 2013. In the period 1970 - 1990, humanity was heavily emitting chemicals that harmed the ozone layer. Until the late 2000s, this harmful emission continued in decreasing amounts. This implies that the study found that the ozone layer would recover in the future. The consensus corroborates this result.

On the topic of the ozone layer, another recent study by Ball et al. talks about a decline in the lower stratospheric ozone that offsets the recovery in the ozone layer [Bal+18]. With more attention paid to other regions of the atmospheric layers, this technique with other time series analysis methods will provide more insight into the underlying behavior of ozone's presence in other regions.

4.2.2 Detrended Fluctuation Analysis

This method quantifies the presence of long-range correlations and self-similarity. Detrended Fluctuation Analysis (DFA) has been used in finance [Kri+10], physical sciences, and for our purposes, in climate-related work as well [Jan+14]. Peng et al. introduced DFA in 1994 [Pen+94]. In this method, the obtained exponent is similar to the Hurst exponent. The key difference between R/S and DFA is that DFA can be used for systems whose underlying dynamics are non-stationary.

The data preparation method is similar to R/S analysis. However, when we divide the data, we do it into various segments with no overlap, referred to as boxes or windows. Each segment will typically represent a local trend within the time series. Then, the local trend is fitted with a polynomial equation. The polynomial is usually linear or quadratic. The estimated local trend is subtracted from the original data points in each segment, leaving residuals or fluctuations behind, called *detrending*. The residuals' root-mean-square (RMS) fluctuations are calculated for each segment. These RMS fluctuation values represent the magnitude of fluctuations. The RMS fluctuations from each segment are averaged; this provides the overall variability in the time series data. Similarly, we perform DFA for different box sizes.

Like R/S analysis, we plot the results on a log-log plot. The slope of the regression analysis, sometimes denoted as α, is the desired exponent. We use $\alpha = 0.5$ to suggest a lack of correlation. Above and below $\alpha = 0.5$ correspond to positive and negative correlations.

5 Predictive Capabilities and Limitations

Nonlinear dynamics has a rich history of prediction strategies based on state-space models. These strategies rely on the reconstruction techniques to predict scalar time series data. One notable example is Lorenz's *Method of Analogues*, which identifies the nearest neighbor in the state-space trajectory and uses its forward path for forecasting [Lor69]. This method works effectively in reconstructed state spaces.

Over the years, with growing interest, researchers have developed various creative strategies for predicting the future behavior of nonlinear dynamical systems. These approaches involve building local models in portions of the reconstructed state space to make predictions. Notably, prediction methods do not always require perfect embeddings to be successful. Even reconstructions that do not meet the theoretical requirements on the embedding dimension can yield accurate predictions, especially when dealing with noisy data. It is essential to avoid over-optimizing predictors to prevent overfitting; when a model fits the training data too closely, it performs poorly on new, unseen data.

Nonlinear time-series analysis in reconstructed state spaces is a powerful approach but has practical limitations. For instance, it typically assumes infinite noise-free observations, which can be problematic in real-world scenarios with statistical properties changing through time, *nonstationarity*, or limited data. The system's level of complexity can be problematic, especially when dealing with systems with large spatial dimensions. Noise effects scale with dimension and filtering and subsequence analysis can help.

The presence of noise in real-world signals is a significant challenge [Cen+00]. Noise can take various forms, including additive random processes and other types of contamination. Understanding the complexity of the underlying process rather than making this distinction is the primary focus in most applications. Filtering noise from nonlinear time-series data is a complex task. Conventional filtering techniques that use specific frequency cutoffs might struggle to handle chaotic signals with wide-ranging frequency patterns effectively. When trying to reduce noise effectively, it is essential to consider the distinct characteristics of nonlinear dynamics, like stable and unstable pathways and attractors' structure or topology.

6 Further Exploration

6.1 Some Base Papers

For any reader who wants to start their journey into nonlinear time series analysis, the following are a few papers that provide a brilliant introduction. There are more base papers than the four listed below; the four below are to be the first step before diving deep into the field.

Nonlinear Time-Series Analysis Revisited
See [BK15]. Bradley and Kantz wrote a concise review in 2015. Kantz was also one of the authors of a book,

Nonlinear Time Series Analysis, with Schreiber in 2004. This paper offers a concise but sufficiently detailed introduction to multiple aspects of nonlinear time series analysis. This paper effectively explains the functions of NTSA and the practical requirements to implement NTSA. We examined this paper in the initial stages of this review. References used in this paper are also robust papers that offer a holistic view of the field.

Deterministic Nonperiodic Flow
See [Lor63]. This 1963 paper fundamentally transformed our understanding of complex systems by introducing the concept of the "butterfly effect," where small initial changes can lead to dramatically different outcomes. Lorenz's work unveiled the deterministic chaos exhibited by certain nonlinear systems, exemplified by the Lorenz attractor, which highlighted non-periodic yet bounded behavior. This groundbreaking paper laid the foundation for chaos theory, demonstrating that deterministic systems can generate seemingly random and unpredictable dynamics, affecting fields from meteorology to physics and economics. This is also a base paper for nonlinear dynamics as a whole.

Geometry from a time series
See [Pac+80]. This paper presents an approach to analyze and understand complex dynamical systems based on the reconstruction of the system's underlying geometry from a single time series of data points; the foundation for what is now known as the method of "attractor reconstruction" or "phase space reconstruction." The authors demonstrate how to recover the geometric properties of a chaotic system, including its attractor and dimensionality, from observational data, enabling the characterization and prediction of complex, nonlinear behavior. This paper is essential for anyone interested in chaos theory, nonlinear dynamics, and the practical applications of these concepts in fields ranging from physics to biology and engineering. It provides valuable insights into the geometric interpretation of time series data.

Determining embedding dimension for phase-space reconstruction using a geometrical construction
See [KBA92a]. This paper presents a novel geometric approach to tackle the crucial problem of estimating the appropriate embedding dimension when reconstructing phase space from time series data. Reading this paper also results in a better understanding of how dimensions come into play in their role in time series analysis.

6.2 Current Research

Machine Learning and Neural Networks
Machine learning, particularly neural networks, has emerged as a powerful nonlinear time series analysis tool. Neural networks, with their ability to capture complex relationships in data, have been instrumental in modeling and forecasting nonlinear time series. They excel in extracting intricate patterns and dependencies that might be challenging to capture with traditional statistical methods. Recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks, particularly, have demonstrated exceptional performance in handling time-dependent data [Don+10; Lin+21]. Researchers leverage these neural architectures to predict chaotic behavior, recognize hidden dynamics, and analyze complex systems in diverse fields from finance to climate science. The adaptability of machine learning and neural networks to nonlinear time series analysis expands our understanding of complex temporal data, enabling more accurate predictions and valuable insights into dynamic systems. With the rise of research in artificial intelligence techniques, models integrating various methods will also gain popularity.

Network Theory
By representing time series data as nodes in a network and connecting them based on specific relationships or dependencies, network theory allows researchers to uncover hidden patterns and structural characteristics in dynamic systems. This approach is used to understand complex interactions, synchronization, and emergent behaviors within various domains including biology, finance, and social sciences. Researchers employ tools from network theory to model and analyze the underlying structure of time series data, leading to insights into system dynamics, identifying critical nodes or elements, and predicting potential disruptions or critical transitions. Integrating network theory with nonlinear time series analysis provides a robust framework for studying dynamic data's intricate and often hidden interdependencies, facilitating a deeper understanding of complex systems and their behaviors. An elaborate work on network approaches to nonlinear time series analysis is by Zou et al. in [Zou+19].

6.3 Some Limitations to Explore

Noise Filtering
Advancing research to mitigate noise from nonlinear time series data is a critical endeavor with broad im-

plications. Noise from measurement errors, environmental factors, or inherent system variability can obscure meaningful patterns and distort the understanding of complex systems. In today's data-driven world, where information is critical to decision-making, extracting reliable insights from complex, noisy data is paramount. Further research into various techniques to decrease noise's impact would benefit NTSA and other fields that heavily rely on data. A 1993 survey on noise reduction methods by Kostelich and Schreiber is in [KS93]. Since then, a significant synthesis or survey has not been done. A 1991 paper on state-space reconstruction in the presence of noise by Casdagli et al. is in [Cas+91].

Dimension Estimating

Another area to explore further is the development of robust and scalable techniques for dimensionality reduction and feature selection. As datasets grow in size and complexity, there is an increasing need for methods to distill the most relevant information from the data efficiently. Researchers should focus on refining existing techniques and creating innovative algorithms to address the dimensionality challenge, as this directly affects the efficiency and interpretability of models in various fields.

Some current, widely-used methods include the Sequential Monte Carlo (SMC) [Kan+09] and geometrical construction [KBA92b].

Lack of Climate Science Application

Nonlinear time series analysis has predominantly found application in finance, such as currency exchange rates and stock price forecasting, but its under-explored potential in climate science holds the promise of significant progress. While it has revolutionized financial prediction by capturing intricate market behaviors, its adoption in climate science remains limited. We can obtain similar results and accuracy in climate science by applying nonlinear time series analysis to climate data.

While compiling literature for this review, it was challenging to identify recent, groundbreaking nonlinear time series analysis models in climate variability compared to their counterparts in finance. We believe that, with similar experiments and studies in climate science, we will learn more correlations between data that we might not currently know.

7 Conclusion and Discussion

Nonlinear time series analysis (NTSA) has been an informal concept for as long as humans have contemplated events unfolding over time. However, it is only recently formalized as a field of study, and even more so, in climate science. This review has navigated the intricacies of NTSA and its applications in climate variability, emphasizing its pivotal role in understanding complex climate systems.

What this review has highlighted is the significant presence of nonlinearity in climate dynamics, which challenges traditional linear models. The mathematical foundations of chaos theory, fractals, and Lyapunov exponents equip researchers with tools to quantify chaos and fathom climate data's self-similarity and scaling properties. Techniques such as state-space reconstructions and delay coordinate embedding prove invaluable in extracting insights from climate time series data.

Looking ahead, the future of NTSA appears promising, with potential integration with machine learning and improved data visualization. However, it is not without challenges, including data quality, computational complexity, and model interpretability, which remain crucial considerations. Collaboration across diverse fields promises innovation and practical solutions for real-world issues. NTSA's prominence, having proven its worth in finance-related industries, is set to expand further, offering valuable methodologies for climate-related research.

The enduring significance of nonlinear time series analysis resides in its core principle of interpreting data's temporal evolution. NTSA's ability to distill complex data into its essential components unveils concealed information and patterns. Moreover, its capacity to establish connections between these patterns and other datasets reveals relationships that might otherwise remain hidden. This enduring utility guarantees that nonlinear time series analysis will continue to be a valuable tool, contributing to our understanding of complex systems and data.

References

[Hur51] Harold Edwin Hurst. "Long-term storage capacity of reservoirs". In: *Transactions of the American society of civil engineers* 116.1 (1951), pp. 770–799.

[Lor63] Edward N Lorenz. "Deterministic nonperiodic flow". In: *Journal of atmospheric sciences* 20.2 (1963), pp. 130–141.

[Lor69] Edward N Lorenz. "Atmospheric predictability as revealed by naturally occurring analogues". In: *Journal of Atmospheric Sciences* 26.4 (1969), pp. 636–646.

[MA79] Benoit B Mandelbrot and Michael Aizenman. *Fractals: form, chance, and dimension.* 1979.

[Pac+80] Norman H Packard et al. "Geometry from a time series". In: *Physical review letters* 45.9 (1980), p. 712.

[Wol+85] Alan Wolf et al. "Determining Lyapunov exponents from a time series". In: *Physica D: nonlinear phenomena* 16.3 (1985), pp. 285–317.

[Bar+88] Michael F Barnsley et al. *Fractals in nature: from characterization to simulation.* Springer, 1988.

[TE89] AA Tsonis and JB Elsner. "Chaos, strange attractors, and weather". In: *Bulletin of the American Meteorological Society* 70.1 (1989), pp. 14–23.

[VG89] Robert Vautard and Michael Ghil. "Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series". In: *Physica D: Nonlinear Phenomena* 35.3 (1989), pp. 395–424.

[Cas+91] Martin Casdagli et al. "State space reconstruction in the presence of noise". In: *Physica D: Nonlinear Phenomena* 51.1-3 (1991), pp. 52–98.

[Sre91] KR1090334 Sreenivasan. "Fractals and multifractals in fluid turbulence". In: *Annual review of fluid mechanics* 23.1 (1991), pp. 539–604.

[KBA92a] Matthew B Kennel, Reggie Brown, and Henry DI Abarbanel. "Determining embedding dimension for phase-space reconstruction using a geometrical construction". In: *Physical review A* 45.6 (1992), p. 3403.

[KBA92b] Matthew B Kennel, Reggie Brown, and Henry DI Abarbanel. "Determining embedding dimension for phase-space reconstruction using a geometrical construction". In: *Physical review A* 45.6 (1992), p. 3403.

[TA92] Timo Terasvirta and Heather M Anderson. "Characterizing nonlinearities in business cycles using smooth transition autoregressive models". In: *Journal of applied econometrics* 7.S1 (1992), S119–S136.

[The+92] James Theiler et al. "Testing for nonlinearity in time series: the method of surrogate data". In: *Physica D: Nonlinear Phenomena* 58.1-4 (1992), pp. 77–94.

[KS93] Eric J Kostelich and Thomas Schreiber. "Noise reduction in chaotic time-series data: A survey of common methods". In: *Physical Review E* 48.3 (1993), p. 1752.

[RCD93] Michael T Rosenstein, James J Collins, and Carlo J De Luca. "A practical method for calculating largest Lyapunov exponents from small data sets". In: *Physica D: Nonlinear Phenomena* 65.1-2 (1993), pp. 117–134.

[Pen+94] C-K Peng et al. "Mosaic organization of DNA nucleotides". In: *Physical review e* 49.2 (1994), p. 1685.

[Ter94] Timo Teräsvirta. "Specification, estimation, and evaluation of smooth transition autoregressive models". In: *Journal of the american Statistical association* 89.425 (1994), pp. 208–218.

[EKR+95] Jean-Pierre Eckmann, S Oliffson Kamphorst, David Ruelle, et al. "Recurrence plots of dynamical systems". In: *World Scientific Series on Nonlinear Science Series A* 16 (1995), pp. 441–446.

[Bar97] June Barrow-Green. *Poincaré and the three body problem.* 11. American Mathematical Soc., 1997.

[RM97] Carl Rhodes and Manfred Morari. "The false nearest neighbors algorithm: An overview". In: *Computers & Chemical Engineering* 21 (1997), S1149–S1154.

[MVG98] Vincent Moron, R Vautard, and M Ghil. "Trends, interdecadal and interannual oscillations in global sea-surface temperatures". In: *Climate Dynamics* 14 (1998), pp. 545–569.

[Cen+00] M Cencini et al. "Chaos or noise: Difficulties of a distinction". In: *Physical Review E* 62.1 (2000), p. 427.

[Pas+00] Mercedes Pascual et al. "Cholera dynamics and El Nino-southern oscillation". In: *Science* 289.5485 (2000), pp. 1766–1769.

[SS00] Thomas Schreiber and Andreas Schmitz. "Surrogate time series". In: *Physica D: Nonlinear Phenomena* 142.3-4 (2000), pp. 346–382.

[HST01] Anthony D Hall, Joakim Skalin, and Timo Teräsvirta. "A nonlinear time series model of El Nino". In: *Environmental Modelling & Software* 16.2 (2001), pp. 139–146.

[Ghi+02] Michael Ghil et al. "Advanced spectral methods for climatic time series". In: *Reviews of geophysics* 40.1 (2002), pp. 3–1.

[EP04] Alicia M Ellis and Eric Post. "Population response to climate change: linear vs. non-linear modeling approaches". In: *BMC ecology* 4.1 (2004), pp. 1–9.

[Cha05] Richard E. Chandler. "On the use of generalized linear models for interpreting climate variability". In: *Environmetrics* 16 (7 Nov. 2005), pp. 699–715. ISSN: 11804009. DOI: `10.1002/env.731`.

[Din06] Jonathan B Dingwell. "Lyapunov exponents". In: *Wiley encyclopedia of biomedical engineering* (2006).

[Vel+07] Guus JM Velders et al. "The importance of the Montreal Protocol in protecting climate". In: *Proceedings of the National Academy of Sciences* 104.12 (2007), pp. 4814–4819.

[Kan+09] Nicholas Kantas et al. "An overview of sequential Monte Carlo methods for parameter estimation in general state-space models". In: *IFAC Proceedings Volumes* 42.10 (2009), pp. 774–785.

[Don+10] Reik V Donner et al. "Recurrence networks—A novel paradigm for nonlinear time series analysis". In: *New Journal of Physics* 12.3 (2010), p. 033025.

[Kri+10] Ladislav Krištoufek et al. "Rescaled range analysis and detrended fluctuation analysis: Finite sample properties and confidence intervals". In: *Czech Economic Review* 4.03 (2010), pp. 315–329.

[LVR12] JH Lucio, R Valdés, and LR Rodrıguez. "Improvements to surrogate data methods for nonstationary time series". In: *Physical Review E* 85.5 (2012), p. 056202.

[Jan+14] Bulbul Jan et al. "Study of probabilistic modeling of stratospheric ozone fluctuations over Pakistan and China regions". In: *Journal of Atmospheric and Solar-Terrestrial Physics* 109 (2014), pp. 43–47.

[BK15] Elizabeth Bradley and Holger Kantz. "Nonlinear time-series analysis revisited". In: (Mar. 2015). DOI: `10.1063/1.4917289`. URL: `http://arxiv.org/abs/1503.07493%20http://dx.doi.org/10.1063/1.4917289`.

[17] "Sea surface temperature change in the mediterranean sea under climate change: A linear model for simulation of the sea surface temperature up to 2100". In: *Applied Ecology and Environmental Research* 15 (1 2017), pp. 707–716. ISSN: 17850037. DOI: `10.15666/aeer/1501_707716`.

[Bal+18] William T Ball et al. "Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery". In: *Atmospheric Chemistry and Physics* 18.2 (2018), pp. 1379–1394.

[ME18] Glenn R McGregor and Kristie Ebi. "El Niño Southern Oscillation (ENSO) and health: An overview for climate and health researchers". In: *Atmosphere* 9.7 (2018), p. 282.

[Zou+19] Yong Zou et al. "Complex network approaches to nonlinear time series analysis". In: *Physics Reports* 787 (2019), pp. 1–97.

[Lin+21] Benjamin Lindemann et al. "A survey on long short-term memory networks for time series prediction". In: *Procedia CIRP* 99 (2021), pp. 650–655.