PREDICTION METHODS FOR $1/f^{\beta}$ PROCESSES WITH APPLICATION TO THE ANALYSIS OF STRIDE INTERVAL TIME SERIES

by

Etienne Zahnd

B.S. in Electrical and Computer Engineering, ENSEA, 2015

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Etienne Zahnd

It was defended on

July 15, 2016

and approved by

Ervin Sejdić, Ph.D., Assistant Professor, Department of Electrical and Computer Engineering

Murat Akcakaya, Ph.D., Department of Electrical and Computer Engineering

Mao Zhi-Hong, Ph.D., Department of Electrical and Computer Engineering

Thesis Advisor: Ervin Sejdić, Ph.D., Assistant Professor, Department of Electrical and Computer Engineering

PREDICTION METHODS FOR $1/f^{\beta}$ PROCESSES WITH APPLICATION TO THE ANALYSIS OF STRIDE INTERVAL TIME SERIES

Etienne Zahnd, M.S.

University of Pittsburgh, 2016

The power law in the frequency spectrum $S(f) = 1/f^{\beta}$ allows for a good representation of the various time evolutions and complex interactions of many physiological processes. The spectral exponent β can be interpreted as the degree of fractal characteristic which in turn makes it some sort of biomarker that gives an idea of the relative health of an individual. This thesis presents a thorough investigation of prediction of the fractal nature of the process with specific consideration given to experimentally measured gait stride interval time series. The goal is to consider the accuracy of several time series prediction methods such as the neural networks, regression trees and bagged regression trees learning method. To test these methods we simulated stride intervals time series as $1/f^{\beta}$ processes. This investigation is to complement previous analyses on predicting the process with which this study compared. It was shown as result of the research that the greatest number of points one can accurately predict is between five and fifteen using the regression tree,the feedforward neural network and the AR model.

Keywords: prediction, stride interval, time series, $1/f^{\beta}$ process, human gait, neural networks, regression trees, autoregressive model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 OVERVIEW OF THE HUMAN GAIT

Postural control is no longer considered simply a summation of static reflexes but, rather, a complex skill based on the interaction of dynamic sensorimotor processes [1]. Human walking, or gait, is heavily influenced by external stimuli and it is controlled and regulated by physiological systems and parameters. One would simply describe the human gait as being the continuous motion of the lower body segments in a repetitive and periodic manner as to not lose balance. That is the concept everyone experiences on a macro scale although there is much more to human locomotion than simply moving one's legs since it is generated thanks to a complex neurological wiring system [1]. Regardless, the following comments on the human gait is a simplified overview of the various phases of an entire completed cycle of the physical process on the macro scale. Simplifying this process allows to easily define the gait process by dividing the cycle in different functional phases which would in turn allow to easily pin point a specific physiological or physical task for analysis purposes [2, 3]. Each of these phases allow for an enhanced visualization of one completed gait cycle, further allowing to distinguish the various processes in motion such as control, feedback and external mechanisms. In this case, the gait process starts off with an initial contact phase where the heel of the foot lands on a surface and ends with the same foot doing the same contact phase and thus another cycle instantly begins. This entire process is human locomotion briefly explained [2–5]. The cycle of the gait process can be broken into eight individual phases, an overview of these (a)-(h) can be found below and it is illustrated in Figure 1.1.

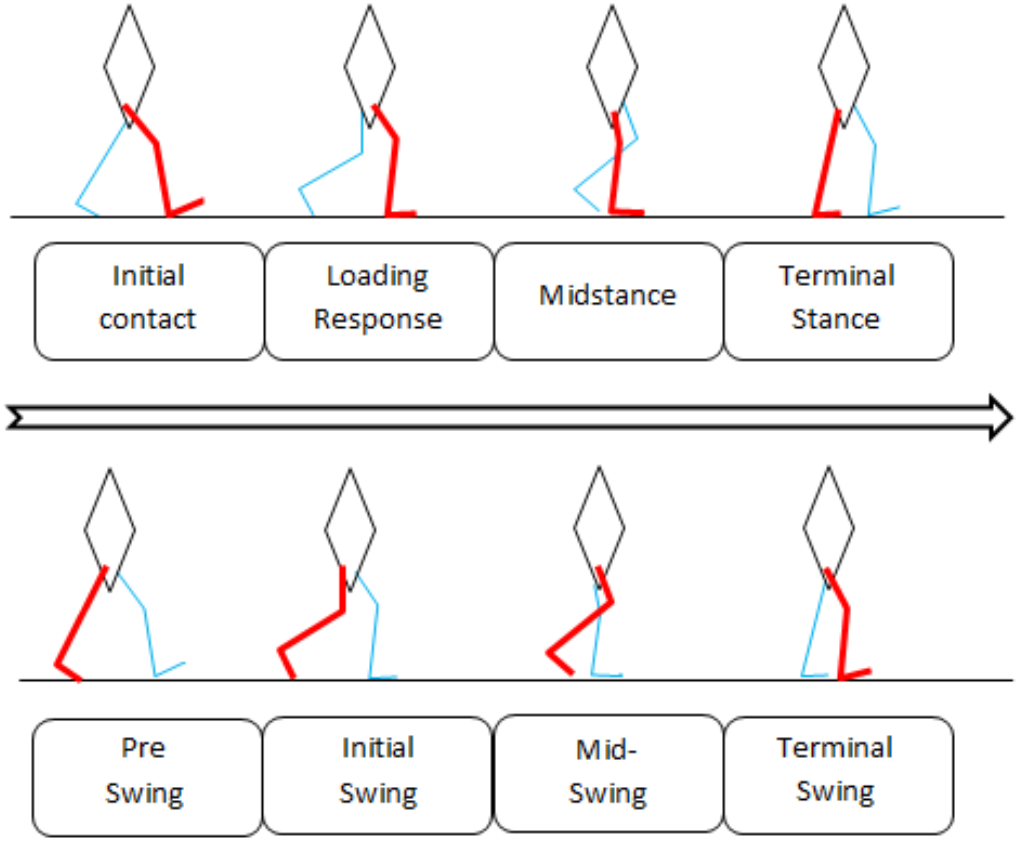


Figure 1.1: Human Walking Kinematics. The thick red line represents the leg which corresponds to the phase below, the blue line is the other leg

(a) Initial Contact: The initial contact phase when an individual's foot hits the floor with the heel. Once in this position, the limb is ready to take on the force that will be applied to it, resulting in the subsequent phase [1–3].

(b) Loading Response: The Loading Response phase begins after the foot of an individual in motion is initially applied on a surface and it ends when the other foot is lifted from the floor. In this state, the first limb flexes in order to absorb the shock so that it can maintain maximum surface contact between it and the surface [1–3].

(c) Midstance: This is the first phase where the foot that made the initial contact is in full support of the body of the individual and the other leg is swinging from back to front. The first foot is still in full contact with the surface and the knee is extending [1–3].

(d) Terminal Stance: In the Terminal Stance phase, the first leg is still supporting the body and it is the last phase where the foot is in full contact with the floor. The body moves forward and the leg is now placed behind the individual [1–3].

(e) Pre-swing: The Pre-swing phase is the last phase where the foot that made the initial contact is still in contact with the surface. In this stance, only the toes are in contact in order to prepare for the swinging phase [1–3].

(f) Initial Swing: The Initial Swing phase is similar to the loading response phase except that the feet state are inverted. The leg swing begins with the initial swing, the first of three phases, where the leg is still behind the individual. The toes leave the surface rendering the leg airborne, ideal for swinging. This allows to carry the body weight forward, continuing with the motion[1–3].

(g) Mid-swing: The Mid-swing phase, as its name suggests, is the phase where the leg is in the middle of the swing period. The knee remains flexed however the hip carries the leg through the end of the swing [1–3].

(h) Terminal Swing: The Terminal Swing phase is the last phase of the swing period and the last one of the entire cycle of the gait process. It ends when the foot hits the floor again. The limb is now in front of the individual, the knee fully extends and now the body is in position to reiterate the whole cycle again [1–3].

As previously mentioned, the separation of the entire gait process into eight phases described above allows a more clear identification of each phase's specific functional objectives and the unique synergistic motions of all the physical components [1–5]. Past studies conducted relating to the gait process may have considered one or more specific events which may occur between or within each of the phases described hereinabove, mainly in a non-macro scale [1,2]. The issue with the eight phase cycle of the gait process given by VT Inman is that it is not entirely precise and not describe every step for any case possible [3]. It was mentioned that the first step of the cycle is defined by the heel of an individual touching a surface [3] however this does not concern every member of the population. Some may have a dysfunctional gait where their heel may never enter into contact with a surface, or may do so during the cycle when the person falls and catches himself during a swinging period for example and the foot hits the ground [2,4,5]. VT Inman introduced the eight phase process

as a way to resolve any misunderstandings that dysfunctional gaits may present to analysis of unique abnormalities [3]. In the following section, an overview of the decline in human gait will be done where the causes and their effects on locomotion will be discussed.

1.2 DECLINE IN GAIT

The human body is comprised of many physiological systems which interact in a nonlinear manner [6–14]. As a result, changes in functional outcomes in a given physiological system may be caused by events in one or more systems in the body [6, 7, 15]. There exist many possible causes that affect the gait process: aging, diseases, genetic disorders and trauma to name a few. These causes have various affects, ranging from benign to significant [4, 5, 16–18]. That complex system is the locomotor system, which influences directly the human gait process. It is consists of many physiological systems, mainly the central nervous system mainly [4, 5]. Without going to much into detail, the locomotor system is itself composed of the cerebellum, the motor cortex, the basal ganglia and various sensors [4, 5]. The locomotor system is very much similar to a control system and can thus be compared to one. Indeed, the cerebellum and basal ganglia receive information for processing and send signals via the motor cortex. Current state information and feedback are provided by internal and external inputs different physiological sensors [4, 5, 7]. When an individual has a healthy gait, a stable walking pattern is drawn and maintained thanks to the functionality of all the components that compose the locomotor system.

As mentioned hereinabove, there exist elements that may alter the locomotor system's ability to correctly and accurately react to changes in the gait process [18]. Hausdorff et al mentions that these neurophysiological changes include for example decreased nerve conduction velocity, decreased proprioception, muscle strength, loss of motor neurons, and central processing capabilities which are notable declines due to advancing age and are therefore natural causes that affect many seniors [18]. However, aging is not the only cause of declining gait, neurodegenerative diseases affect a wider range of people in the age spectrum. Amyotrophic Lateral Sclerosis (ALS) is one of them. It significantly affects the function of

the motor neurons of the cerebral cortex, brain stem, and spinal cord [17]. Muscle weakness, increased fatigue and decreased endurance are characteristic of ALS [19–21]. These changes in the gait process directly influence the walking velocity by significantly reducing it and therefore increasing the stride interval time. Other neurodegenerative diseases that affect the locomotor system, more precisely in the basal ganglia, in a negative way are Huntington's disease (HD) and Parkinson's disease (PD) [18]. Both Huntington's and Parkinson's disease are known to greatly affect the central motor control where one of the outcomes can be visible on the macro scale which is an intensive shaking of the limbs [18, 22]. Although these diseases have different effects on the human locomotor system and thus the gait process, all three increase the stride interval time since movement is decreased [18, 22]. However it is important to note that the increase in stride interval time does not imply that one suffers from a neurodegenerative disease, therefore the fluctuations of the stride interval must be considered in order to make such conclusions [17, 18]. That being said, it is apparent that once a component of the human locomotor system is affected by change from diseases, aging or even trauma, for example a stroke, the consequence can be an abnormal gait. However, the identity and severity of the underlying mechanism(s) causing the functional decline are still unknown, and can be extremely difficult to identify and characterize due to the highly nonlinear and complex interactions of the constituent physiological systems [17, 18, 23, 24].

1.3 CURRENT METHODS TO DETECT FUNCTIONAL DECLINE IN GAIT

Analysis of the gait process is extremely challenging. Given the high number of time variant physiological systems and their according parameters, a complete quantitative perspective of the gait process has very high dimensionality [25]. Potential causes are parameters of some general, neural and physical processes that may consist of the kinematic and kinetic data of joint angles, or even velocities, not to mention some anthropomorphic parameters [2, 25]. In conclusion to these important amounts of critical variables, alternatives methods that reduce the amount of data yet keep the key information are necessary [2, 25].

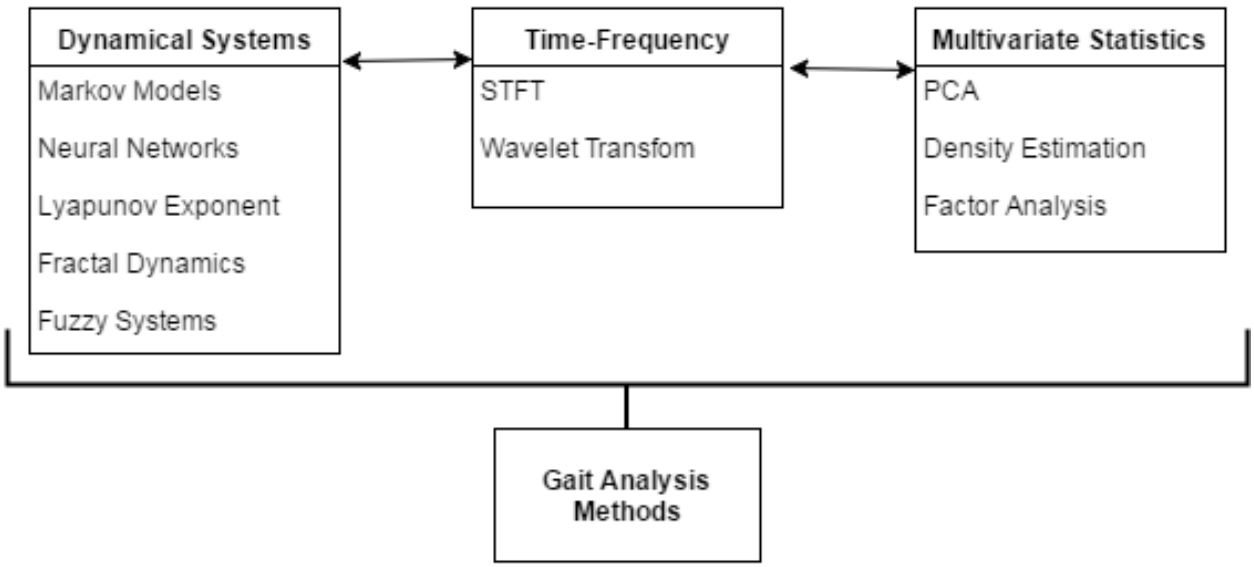


Figure 1.2: Gait Analysis Methods

Physiologists recognize the need for effective methods that reduce and extract useful information from highly correlated time-dependent gait data [4, 5, 25–27]. Such methods would greatly expand the understanding of motor control mechanisms and have immense benefit for monitoring disease progression and applying appropriate therapies [17, 25, 26]. Numerous investigations were achieved regarding various approaches that would allow to analyze the gait process. Figure 1.2 overviews some of these methods that expands toward a large set of various fields such as engineering, physics, computer science to name a few [6, 9, 10, 25, 26, 28–30].

With such a wide variety of available analytical tools, the determination of an appropriate method to investigate health and decline in gait data is also difficult. In many cases, the choice of an analytical method is generally dictated by the nature of the available data. The main challenges of the gait analysis problems these methods aim to address are the high dimensionality, variability, temporal dependence, and nonlinear relationships of gait data [25]. The following explanation will be a brief discussion of Figure 1.2. Multivariate statistical methods such as the factor analysis, the principal component analysis and the density estimation are generally useful when addressing the problem of high dimensionality and variability as they have been shown to provide helpful insight into interpretations of gait data [25]. Unfortunately, these multivariate statistical methods have limited applications due

to the supposedly linearity between gait variables [25]. Furthermore, when dealing with five or more variables, these statistical methods are severely limited due to the assumption of linearity [25]. Moreover, when applying the same tests with multivariate statistical methods while using different patients, tools and experimental design, a discrepancy between these tests is noticeable because of these differences in experimental design [2, 25]. Therefore, the high variability of statistical measures between various studies makes statistical conclusions drawn from gait date more difficult [25]. Finally, in physiology the time evolution of variables and systems in a considered process is critical. Generally, model parametrizations of the time evolution of parameters assumes that there is stationarity. This implies that patterns in the gait process that are sensitive to time will not be picked up by these models [7, 9, 10, 25, 30]. On the other hand, dynamical system models such as the Markov models have been developed in a way that allows to describe the gait process under random perturbations, however it cannot reveal any underlying mechanism that governs the system's behavior [7, 27, 31]. That is because these models assume stationarity, this leads to time-sensitive patterns present in the gait process to not be picked up [7, 9, 10]. The nonlinear relationships between these gait parameters is a natural observation of the dynamics of systems, but such relationships defy traditional analytical description [7, 25].

There is therefore a need to find alternative methods in order to address the complexity and behavior of the variability of the gait data. These alternative methods include some approaches such as artificial neural networks, fractal analysis and fuzzy analysis. Artificial neural networks have been shown to be a promising alternative tool, however their use are quite complicated and their result are hard to understand and quantify [25]. Similarly, fractal and fuzzy analysis have also been shown to be promising tools in a clinical setting [9, 10, 25, 30]. The difference between these two is that fuzzy analysis treats variability as a non-probabilistic uncertainty, whereas fractal analysis considers variability to be correlated and the outcome of long range dynamics [7, 9, 10, 25, 30].

Nowadays there exist a variety of ways where stride intervals studies are used to detect gait functional decline. Indeed, as previously mentioned, there exist a correlation between the latter and the stride interval time series [17, 18]. Recent studies have shown that a

slow gait does not necessarily imply that an individual presents any cognitive declines but is rather an early indicator of a pathology that may or will affect the locomotor system [18, 32–35]. Mielke et al has shown that a faster gait speed was usually associated with a better performance in memory, global cognition, and executive function [32]. At baseline, it was hence usually associated with less cognitive decline across all domain-specific and global scores whereas for a slower gait, both gait speed and cognitive scores declined over time [32]. In order to observe these results, the notion of stride interval was considered. This is defined as the time between subsequent initial contact phases, and its value is considered as the output of the locomotor system [2–5, 16–18, 25]. The understanding of physiological processes as scale invariant parameter fluctuations provides a simple distinction between healthy and pathological processes [16–18]. Using some of these dynamical systems invoked earlier, the investigation of the application of dynamical system as a way to predict gait stride intervals can provide meaningful insight into the presence of physiological decline. Therefore, there is a need to find a method that would allow to measure prediction of the gait process as a way to give an early warning on the decline of an individual's gait.

1.4 RESEARCH OBJECTIVE

Fractal analysis techniques is a promising tool for developing computational biomarkers of decline due to aging, disease, and trauma. This work seeks to further solidify the foundation of fractal analysis as a medical diagnostic tool by using a fractal process known as the $1/f^{\beta}$ process which has been shown to accurately simulate stride interval time series, and to make clearer a direction to progress toward the long term goal of developing machine learning algorithms applied to the medical world. Reviews of the current techniques have indicated that promising alternative methods such as fractal analysis, fuzzy clustering, and artificial neural networks still require additional inquiry to clarify their potential as a phsyiological analysis tool, and here to the application of gait analysis. Fractal analysis shows significant promise for providing insight into gait processes that is not attainable with the previously mentioned traditional methods. The clinical applicability and accessibility of fractal time

series analysis here is an important goal. To these ends, this work mainly aims to quantify the performance of fractal time series analysis under ideal and constrained conditions, and to demonstrate the utility of these methods to signals derived in a physiological setting.

The first goal of this thesis will be to carry out a comparative analysis of machine learning algorithms to determine the best possible method available that would accurately model the existing $1/f^{\beta}$ process. Four different methods will be compared, two neural networks - the feedforward neural network (FFNN) and the layer recurrent neural network (LRNN) - and two regression trees - the regression tree (RT) and the bagged regression tree (BRT). This will be done in order to visualize the accuracy of these methods.

The second aim will be to set a number of points to forecast and actually determine until which point prediction remains accurate. This is done by extrapolating data using ARIMA models with a given seasonality that allows for a forecast of the $1/f^{\beta}$ power law. The number of forecasted points chosen in this study were five, fifteen, twenty-five and thirty-five. The tests will include the entire range of the $1/f^{\beta}$ process with a total number of strides being equal to three hundred which is determined to be the average number of strides done by a healthy subject in five minutes even though a common limitation in acquiring physiological data, such as gait stride intervals, is the time series length [7,29,36]. However this thesis will disregard this since prediction will be done on a wide range of β value using simulated data from the $1/f^{\beta}$ itself. It has also been recognized that the parameters of many physiological processes, such as stride interval time series, are by nature not zero mean [4,5,16–18].

Chapter 2 provides background information regarding the elements that are discussed in this thesis. Then Chapter 3 describes the methodology for the different phases of the research. Chapter 4 presents the prediction results and the final selected features, as well as the relevance of these prediction methods. Implications for these results are discussed in Chapter 5. Finally, conclusions and future directions for the study are provided in Chapter 6. Chapter 7 acknowledges the support of those who helped the author in completing the study, followed by a list of references.

2.0 BACKGROUND

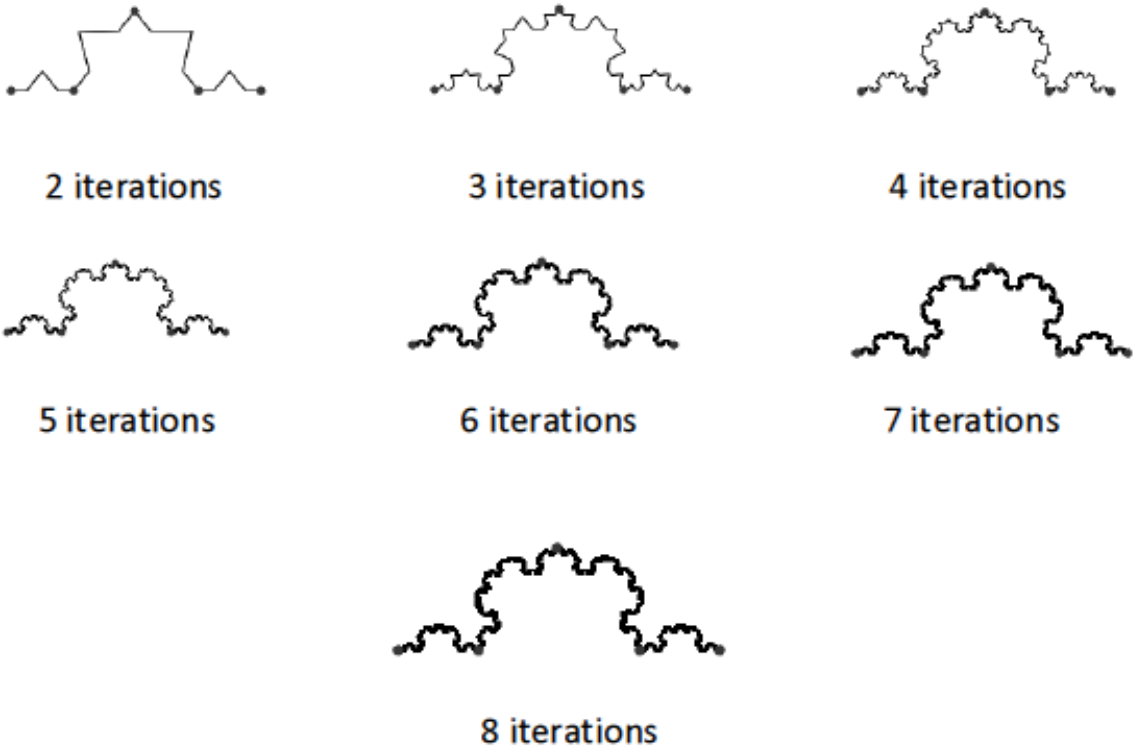
2.1 FRACTAL TIME SERIES ANALYSIS

2.1.1 Fractals in Mathematics

Originally conceived for geometric purposes, the mathematical concept of fractals was declared to explore the existence of scale independent features [7, 37–39]. Generally, complex fractal geometries are created by applying a basic rule of transformation, otherwise known as the generator to a simple geometric object such as a line, triangle, or square, that is usually called the initiator [7, 39] A very well-known example is the Von-Kock curve, also known as the "snowflake" fractal. It is easily recursively constructed by a line initiator. The generator, however, is applied by an equilateral triangle at the midpoint of the line, as can be seen in Figure 2.1. The process is repeated infinitely and that infinite iteration transformation results in a geometry with the identically complex structure at all scale lengths. In the Euclidean geometry, the set of numbers $\mathbb{R}^n$ for $n = 1, 2, 3$, in other words, 1D, 2D or 3D sets of numbers, are used to describe the shape and spacial coordinates of objects in space [7, 37] However, the Euclidean description of inherently complex structures would require a large number of objects and corresponding spacial coordinates to describe the geometry at all scales [7]. This can be seen in Von-Koch's curve fractal by observing it at a shorter scale, as demonstrated in Figure 2.2. This comes to show that at lower dimensions, when the Euclidean coordinates are in $\mathbb{R}^n$ for $n = 1, 2$, the Von-Koch curve fractal cannot be sufficiently described on all scales [7, 37–39]. A direct result of this recursively defined structure is that spaces of integer dimension are inadequate for completely representing it. Additionally, a set of traditional geometric axioms inherently cannot describe the outcome of the structure

(a) Initiator

(b) Generator



(c) Resulting Fractals according to the number of iterations

Figure 2.1: Von-Koch snowflake fractal

in the same way the recursive algorithm can, which contains underlying mechanism of the outcome [7,38,39]. To be able to fit the infinite scales of existence of the geometrical features of these fractals, spaces of non-integer dimension $\mathbb{R}^{n+m}$ where $0 < m < 1$ have to be used [7, 37–39]. Since m is a fractional number, $m \in \mathbb{D}$, dimensions such as $\mathbb{R}^{n+m}$ are fractional, commonly referred to as fractals, hence the name. In the instance of the Von-Koch curve, the complex structure is given simply in terms of an initiator and generator. The general utility of fractal space then is a unique description of complexity by a simple recursive algorithm that captures the essence of the object. This is used in favor of whole integer space

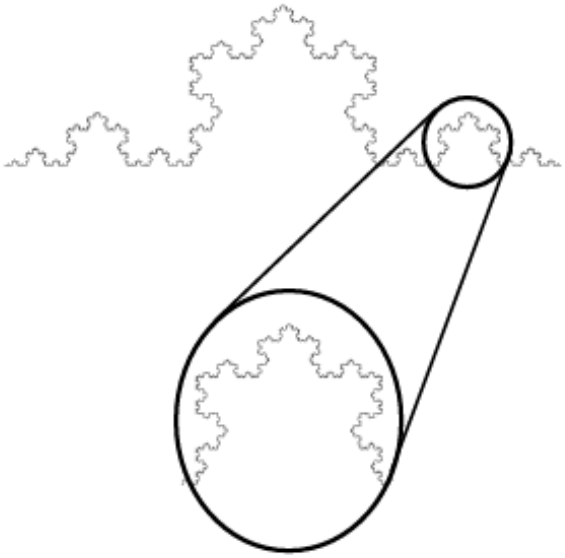


Figure 2.2: Scale invariant feature of the Von-Koch snowflake fractal

and Euclidean geometry requiring a large number of coordinates and axioms to be able to describe an very complex process [7].

Fractals that are mathematically generated by a recursive algorithm display the same feature over an infinite range. Fractals can occur naturally, the perfect example is the snowflake as it is considered to have a fractal-like shape however they can also be statistical. Both exhibit feature scaling over a limited range. An Example of an exact fractal is the Mandelbrot Tree, as seen in Figure 2.3 [7]. The infinite recursive iteration in the Mandelbrot Tree creates an identical branching feature which continues infinitely at all smaller and larger scale lengths. Given this definition, it is possible to consider some structures in day to day lives to be fractals, a couple of examples are the neural and arterial networks, although the branching is not as exact as the Mandelbrot Tree of Von Koch's snowflake and the scale length is not infinite as well. However, this is not to say that the properties they exhibit are not striking.

The measurement of a feature in a fractal object does not converge to any value on a single scale. There exist a parameter q which has a power law scaling relationship instead.

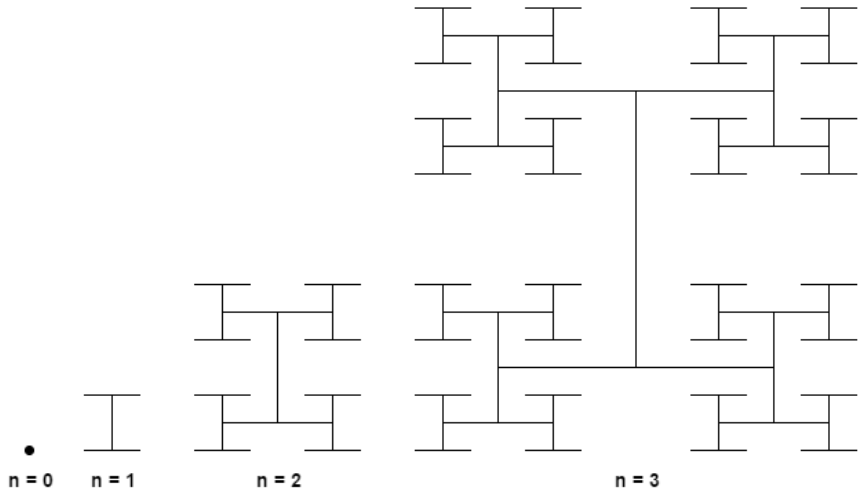


Figure 2.3: Mandelbrot tree with varying number of iterations

This relationship is given in Equation 2.1 [7] :

$$q = ps^{\varepsilon} \tag{2.1}$$

where s is the scale, p is the factor of proportionality and ε is the scaling exponent. As with all exponents, the scaling exponent ε is easily retrievable by applying the *log* function on the previous equation. The value is therefore:

$$\varepsilon = \frac{\log q - \log p}{\log s} \tag{2.2}$$

It is critical to be aware that the range of scales s is restricted for natural physical structures and phenomena.

2.1.2 Fractal Time Series

The time series of many physiological processes' parameters such as arterial blood pressure, breathing, heart rate, and walking have been observed to possess various complex statistical properties [9,10,14,18,24,40–42]. This phenomenon is due to the time evolution and complex interactions of many dynamical systems, imposed with random fluctuations, resulting in chaotic processes [43]. Some descriptions of the statistical qualities of these processes may be of a slow exponentially decayed autocorrelation function, a $1/f$ type power spectral

density, or a heavy-tailed probability distribution function [37, 44–46]. In such processes, it is possible to consider the time evolution of variables represented by a time series $x(t)$. The process has the property of being long range dependent when the autocorrelation function $r_{xx}(\tau)$ of $x(t)$ satisfies the following condition:

$$\int_0^\infty r_{xx}(\tau)d\tau = \infty \tag{2.3}$$

This means that the autocorrelation function is not integrable [37, 47]. This implies that the nonintegrable autocorrelation function can be expressed as follows:

$$r_{xx}(\tau) \approx c|\tau|^{-\beta} \qquad (\tau \to \infty) \tag{2.4}$$

where $c > 0$ and $0 < \beta < 1$. This relationship implies a power law and long range dependence in the autocorrelation of $x(t)$ [37]. Similarly, the autocorrelation function reveals special properties of the power spectral density $S(\omega)$ of $x(t)$:

$$S(\omega) = \int_0^\infty r_{xx}(t)e^{-j\omega t}dt \tag{2.5}$$

In the case of a long range dependent time series, $S(\omega)$ does not exist in the above form [37]. However, when the autocorrelation function approximated by the power law $c|\tau|^{-\beta}$, the power spectral density can be found by:

$$F(|\tau|^{-\beta}) = 2sin(\frac{\pi\beta}{2})\Gamma(1-\beta)|\omega|^{1-\beta} \tag{2.6}$$

where F is the Fourier transform and Γ is the Gamma function [37, 48, 49]. It follows that the power spectral density of the long range dependent process also possesses the power law property $S(f) = 1/f^\beta$ [37, 39, 42, 50].

This reasoning can be applied to the autocorrelation function where the relationship is given in terms of $p(x)$, the probability distribution function.

$$r_{xx}(\tau) = \int_{-\infty}^\infty x(t)x(t+\tau)p(x)dx \tag{2.7}$$

The integrals that define the mean μ_x and variance σ_x^2, given in Equations 2.8 and 2.9, are non-existent when dealing with a long rang dependent time series $x(t)$.

$$\mu_x = \int_{-\infty}^{\infty} xp(x)dx \tag{2.8}$$

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 p(x)dx \tag{2.9}$$

This implies that the tail of the probability distribution function $p(x)$ is not exponentially bounded, i.e. it is heavy-tailed [37,51]. There are noticeable differences between a time series in the $\mathbb{R}^1$ space and a time series that is long range dependent and possesses the power law type autocorrelations, a $1/f$ power spectral density and a heavy-tailed probability distribution function. The latter is fractal and it exist in the $\mathbb{R}^{d+1}$ where $d \in [0, 1]$ [37]. The existence of these properties are indicative of other unique characteristics which can reveal fractal behavioral mechanisms of the process beneath the apparent chaos [30, 37, 52]. However, the long range dependence of time series like these and absence of the mean and variance such measures are not enough to describe the statistical nature of these properties. This lead to the introduction of the Hurst exponent H for cases like these and it is implemented similarly to the long range dependent autocorrelation equation given in Equation 2.4:

$$r_{xx}(\tau) \approx c|\tau|^{2H-1} \qquad (\tau \to \infty) \tag{2.10}$$

Unlike the geometrical fractals, fractal time series cannot be self similar because the dimensions of a time series, which is the amplitude over time, are different. This implies that any power law exhibits the property of self affinity [7]. In general, mathematical and statistical fractal geometries and structures are self-similar, whereas amplitude versus time processes are self-affine [7]. However, there is no importance in distinguishing the differences between self-similar and self-affine in this thesis. Generally the description of power law behavior of parameters is referred to as self similarity.

In this study, the it was assumed that the scaling index β is time invariant which would imply that the time series is monofractal. The latter is a simplified variant of the multifractal time series [7, 53, 54]

2.1.3 Obtaining the Spectral Exponent β

It has been previously defined that the power spectral density is an informative perspective of fractal processes, which shows inverse power law scaling behavior by $S(f) = 1/f^{\beta}$. These types of processes are commonly referred to as $1/f^{\beta}$ processes [7,26,42,50,55–57]. Generally, these $1/f^{\beta}$ processes can be classified, it can either be the fractional Gaussian noise class (fGn) or the fractional Brownian motion class (fBm) [7, 29]. For fGn class signals, the probability distribution of a segment of the signal is independent of the segment size and its temporal position in the signal [7]. Thus, the correlation structure and any statistical descriptions of the process do not change over time, so the process is said to be stationary [29]. In an fBm signal, the probability distribution in a larger segment is equal to a distribution in a smaller segment when the distribution in the large segment is rescaled [7]. The inverse power law relationship is seen for the calculation of a statistical measure m on the segment of length n:

$$log(m_n) = log(p) + Hlog(n) \tag{2.11}$$

where p is a proportionality factor and H is the Hurst exponent. The latter verifies $H \in [0, 1]$. The Hurst exponent is a commonly used tool for indicating the fractal nature of a fractional Gaussian noise or fractional Brownian motion process [58–60]. These processes have the interesting property that when iterating the summation of the fGn signal, it results in an fBm signal [7]. In the general case, a process, either from the fGn or the fBm class can be classified in the other class by either intergrating or deriving the signals [7,42,50,55,56]. This implies that there is a need for a unique Hurst exponent for each of these classes. Similarly to the signals, these Hurst exponent can be called H_{fGn} and H_{fBm} for the fGn and fBm classes respectively. Since they are both Hurst exponents, their values are inside the $[0, 1]$ interval [7,29]. There is a special case in the Hurst exponent that applies for both classes and it is when $H = 0.5$. When $H_{fGn} = 0.5$, the special case is that the signal is considered to be a white Gaussian noise. $H_{fGn} < 0.5$ is anti-correlated Gaussian noise, whereas $H_{fGn} > 0.5$ is correlated [26]. When $H_{fBm} = 0.5$, the signal is considered to be red noise, or Brownian motion [7, 26]. $H_{fBm} < 0.5$ is anti-persistent Brownian motion, whereas $H_{fBm} > 0.5$ is persistent Brownian motion [7]. White Gaussian noise is the characteristic process of the

fractional Gaussian noise class of $1/f^{\beta}$ processes [7]. A very important property of white Gaussian noise in general is that energy is equally distributed regardless of the frequency. When plotting the power spectrum, the result would be an offsetted flat line implying that $\beta = 0$. As for the fBm class, the Brownian motion, or red noise, is characteristic of that class. The β value for signals in this class is equal to 2 as can be seen in the power spectrum [7, 17]. Contrarily to the fGn case, these types of processes are non-stationary. There is another special case when $H_{fBm} = 0$ in which case the noise is called pink noise and $\beta = 1$ [7]. In that case, correlation exists but it is not strong enough compared to the other cases [29]. An overview of all these cases can be found in Figure 2.4 for when the Hurst exponent $H \in \{0, 0.5, 1\}$.

As demonstrated hereinabove, there is a correlation between the Hurst exponents H_{fGn} and H_{fBm} and the spectral exponent β of the power spectrum of $1/f^{\beta}$. That relationship is given in equations 2.12 and 2.13 [7].

$$H_{fGn} = \frac{\beta + 1}{2} \tag{2.12}$$

$$H_{fBm} = \frac{\beta - 1}{2} \tag{2.13}$$

These equations imply that for a Hurst exponent $H \in [0, 1]$, the value of β is such that $\beta \in [-1, 3]$ and the border between the fGn and the fBm classes lies at $\beta = 1$ [7, 29]. Now that the relationship has been introduced as well as the way of obtaining the β value using the power spectrum, similarly to Figure 2.4, an overview of white Gaussian noise ($\beta = 0$), pink noise ($\beta = 1$) and red noise, or Brownian motion ($\beta = 2$) can be found in Figure 2.5 along with the respective power spectrum plot displaying the slope which is, as previously mentioned, the value of β for that signal.

There are many ways to mathematically obtain the spectral exponent. For a $1/f^{\beta}$, β values can be estimated in time, frequency or time-scale domains. These methods include the bridge detrended scaled window variance (bdSWV) and the detrended fluctuation analysis which are time domain methods, various PSD linear regression methods as seen hereinabove which are frequency domain methods and finally the average wavelet coefficient (AWC) which introduces the continuous wavelet transform (CWT) which is a time-scale domain method.

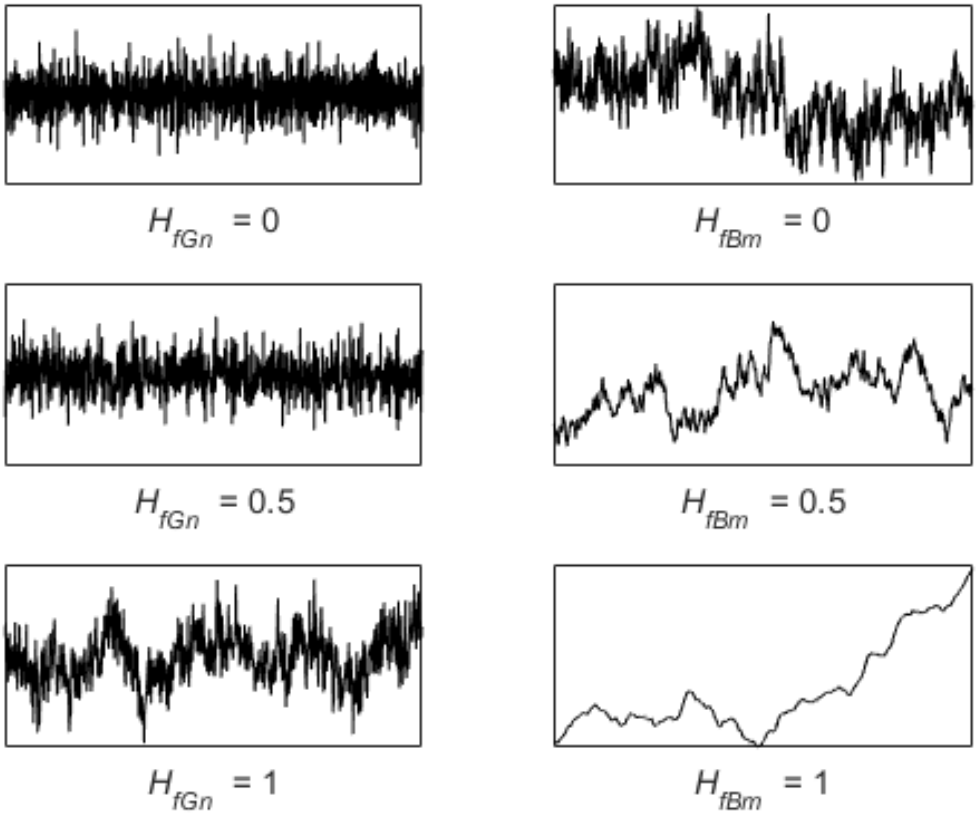


Figure 2.4: Range of the fGn and fBm class signals

A 2014 study that assessed various characterization methods of $1/f^{\beta}$ processes has shown that the spectral exponent β was best estimated when using the time-scale method, the AWC [61]. This method is especially informative, as it provides information of shifted and stretched, or rescaled, versions of the basis mother wavelet function [61, 62]. Moreover, the wavelet transform is not biased for low and high frequency components of the process, and provides precise characterization of these dynamics [7, 61]. The AWC was first introduced by Simonsen in 1998 and it is defined as follows [63]. For a signal $h(x)$, the mother wavelet function is defined as:

$$\psi_{a;b}(x) = \psi(\frac{x-b}{a}) \tag{2.14}$$

where the wavelets chosen are usually the twelfth Daubechies wavelet as given in Simonsen's study [63]. When dealing with wavelet transforms a certain number of levels must be re-

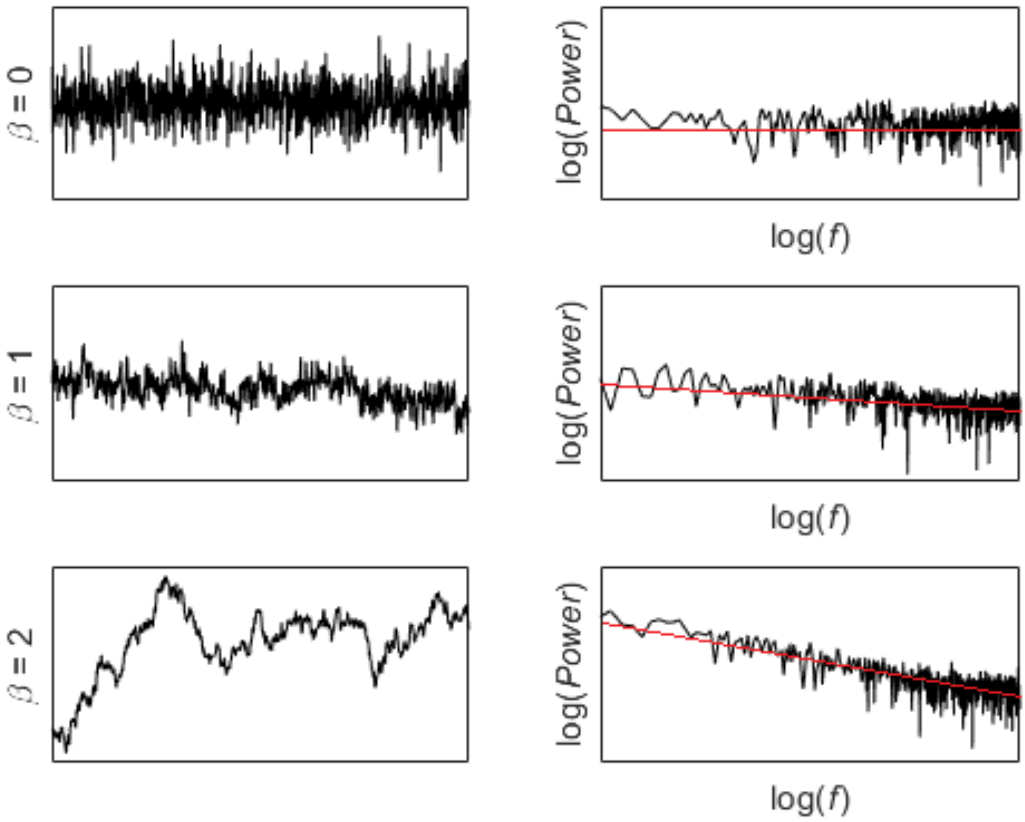


Figure 2.5: Signals of length N for $\beta \in \{0, 1, 2\}$ with respective PSD and linear regression

spected. It is chosen with respect to the signal length, determined as never lower than 2^3 or greater than 2^7 [64]. The mother wavelet is a zero-mean function, where a is the scale coefficient and b is the translation coefficient. Knowing this, the definition of the AWC can be given [61, 63]:

$$W[h](a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi^*_{a;b}(x)h(x)dx \tag{2.15}$$

To find the averaged wavelet coefficient c_n at level n, the arithmetic mean with respect to the translation coefficient b is calculated by:

$$c_n = \langle W[h](a,b) \rangle_b \tag{2.16}$$

The mean is calculated at every level n. The means versus the levels are then plotted on a log-log plot and the slope will then then by $H_{fBm} + \frac{1}{2}$. Therefore $\beta = H_{fBm} + \frac{1}{2}$ [7, 61].

2.1.4 Physical Meaning of Spectral Exponent β

The spectral exponent β of the $1/f^\beta$ process applied to the human gait is considered to be a biomarker as previously mentioned. This means that β is able to quantify the physiological health of an individual regarding his locomotive abilities, also known as gait. Theoretically, a healthy individual would exhibit an index value of $\beta = 1$ whereas an indivual with a pathological gait would rather have a β value that can be equal to any other value while still being between 0 and 2 [7, 14]. However, these results may vary in real life applications, for example a healthy patient may exhibit an index that would be deemed unhealthy but overall, the β values differ depending on the state of a person [4, 5, 7, 14]. A way of predicting the future state of the $1/f^\beta$ process applied to stride interval time-series would have a great impact in the medical field. The issue lies in the number of strides that can be forecasted. This thesis seeks to determine the best method available for forecasting these type of signals by applying machine learning algorithms and extrapolation of data in order to properly forecast future state.

2.2 MACHINE LEARNING ALGORITHMS FOR TIME-SERIES PREDICTION

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory to artificial intelligence. In this study, the $1/f^{\beta}$ process is first modeled using a few machine learning algorithms since it provides an accurate model of the process for extrapolation purposes which will be the second step. Next is an overview of some of literature's most prominent methods. Prior to this, a small definition of supervised learning will be given.

2.2.1 Supervised Machine Learning

Previous studies and breakthroughs in computational algorithms have come to show that it is possible to apply machine learning algorithms to stride interval time series problems. Machine learning tasks can be separated into two major subfields, which depend on the nature of the available dataset. If the pattern recognition system has a labeled training data, that is to say each observation maps to a given class, the learning task is called supervised learning. The other subfield is called unsupervised learning. It is a task that occurs when the dataset is unlabeled. Therefore, the algorithm will have to find the different hidden patterns on its own. Since the $1/f^{\beta}$ process is a stochastic process, supervised learning methods will be used.

As mentioned, supervised machine learning, whose general definition is given in Definition 2.2.1 where the goal of the computer is to map labeled input features to output classes [65, 66]. The idea behind this concept is to create algorithmic tools that are able to perform accurate predictions automatically with the help of past knowledge in order to output future estimations.

Definition 2.2.1. *Supervised Machine Learning*

Given a set of data $S = \{\forall i \in [\![1, N]\!]/(x_i, y_i)\}$ of size N, the supervised learning task consists in finding a correlation between x and y by actively learning. This is achieved by providing a new input x^* and obtaining a new output y^* that is accurate.

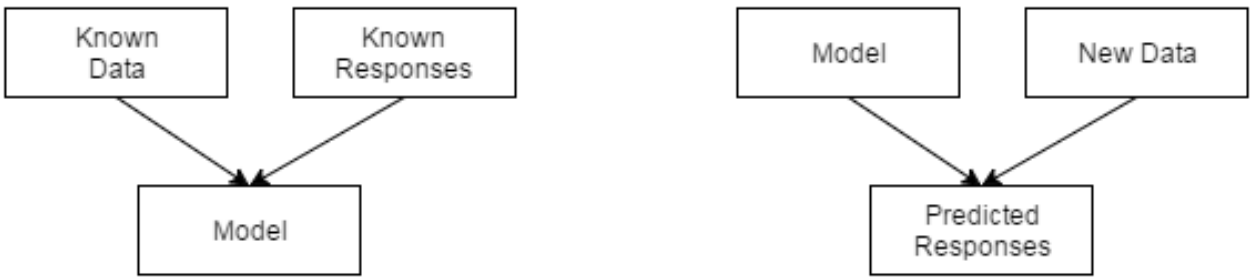


Figure 2.6: General supervised machine learning cycle representation. Left: Step 1. Right: Step 2. Adapted from [77, 78]

The goal of supervised machine learning is to obtain y using the known respective value x. This equates to knowing the value of the conditional probability $P(y|x, S)$. It thus helps to understand why supervised learning is called that way since it refers to a supervisor which is in charge of labeling y for each x in the dataset S.

Supervised machine learning algorithms generally follow a pattern which is given in Figure 2.6. The first step is to acquire a model of the process at hand using known sets. These sets include data and the response of the process when applied to the previous data. It requires either an expert to pick the most descriptive attributes, or a selection methods to segregate the data [67]. Yet, errors can still be present in the dataset, making pre-processing of the data a crucial task in supervised machine learning [68, 69]. This explains the importance of having to know the response of the dataset when processed through the model in order to train it. Once the model is obtained, a new dataset is applied to it. The predicted response of the new data is acquired.

There exist various supervised machine learning methods but can be classified in either one of two categories of algorithm, classification or regression algorithms. Table 2.1 lists a few of them [70–76]. As can be seen, both neural networks and decision trees can be classified in either category.

Table 2.1: List of various supervised machine learning algorithms

Classification Algorithms	**Regression Algorithms**
Neural Networks	Linear Regression
Naive Bayes Classifier	Nonlinear Regression
Decision Trees	Generalized Linear Models
Discriminant Analysis	
Nearest Neighbors (kNN)	
Support Vector Machines (SVM)	

2.2.2 Regression Algorithms Description

For time series prediction, a few methods can be used in supervised machine learning classification or regression category. Regression algorithms are best suited for this task, notably regression trees and neural networks where both include various subgroups [72, 73]. Same goes for linear and nonlinear regression algorithms where a vast number of methods can be found with varying applications [79].

2.2.2.1 Linear and Nonlinear Regression Algorithms .

Table 2.2 is an overview of various linear and nonlinear regression models [80].

Table 2.2: List of linear and nonlinear regression algorithms

Linear Algorithms	**Nonlinear Algorithms**
Simple	Gauss-Newton Algorithm
Multiple	Gradient Descent Algorithm
Multivariate	Levenberg-Marquardt Algorithm

Linear algorithms are can either be simple, multiple or multivariate. They are defined by the general equation:

$$y = \gamma_0 + \Sigma\gamma_i X_i + \epsilon_i \tag{2.17}$$

where γ represent linear parameter estimates , ϵ is the error variable and X_i are the independent variables called predictors. Simple linear regression models are very common, they only deal with one predictor whereas multiple linear regression models deal with multiple predictors as the name suggests. The latter must not be mistaken for the multivariate linear regression models because although the name makes it seem that they are similar, the multivariate one deals with multiple response variables.

On the other hand, nonlinear regression models are the most useful tools in this study because they are best suited for processing stochastic models [80]. Table 2.2 includes a list of the most common algorithms. Generally, a nonlinear regression algorithm is defined by the following equation [81]:

$$y = f(X, \gamma) + \epsilon \tag{2.18}$$

where the nomenclature is the same as for Equation 2.17. Each y is modeled as a random variable with a mean given by a nonlinear function $f(X, \gamma)$ which varies according to the algorithm used. A brief definition of both the Gauss-Newton and Levenberg-Marquardt algorithms is given below.

Gauss-Newton Algorithm - The Gauss-Newton algorithm is similar to Newton's method for finding the minimum of a function. The algorithm is as follows [82]:

$$\mathbf{x}_{n+1} = \mathbf{x}_n + (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\mathbf{r}(\mathbf{f}_n) \tag{2.19}$$

where $\mathbf{J}$ is the Jacobian matrix.

Levenberg-Marquardt Algorithm - The Levenberg-Marquardt algorithm, also known as the Damped Least-Squares method, is an interpolation of the Gauss-Newton algorithm and the gradient descent method however it is quite slower. First introduced by Levenberg and Marquardt, the algorithm that bears their names is often used due to its robustness

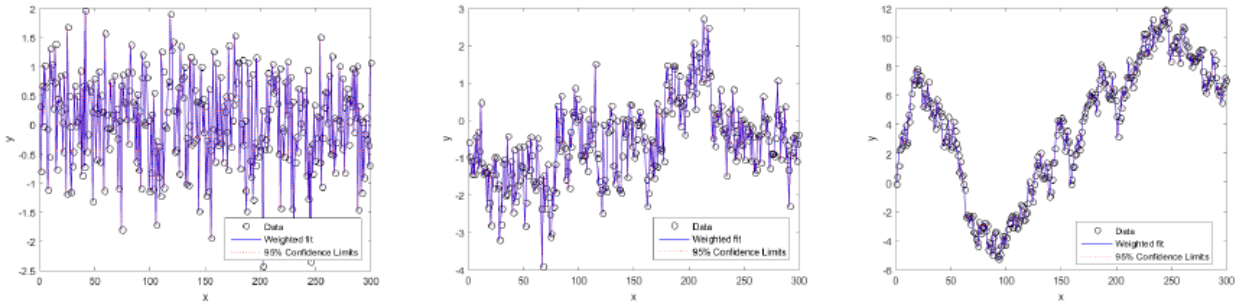


Figure 2.7: Curve fitting $1/f^{\beta}$ process using the Levenberg-Marquardt Algorithm. Signals: o is the data generated by the stochastic process, - is the fitted curve. Left: $\beta = 0$, middle: $\beta = 1$, right: $\beta = 2$.

compared to the Gauss-Newton algorithm, meaning that a solution will eventually be found. The Levenberg-Marquardt algorithm is given as follows [83, 84]:

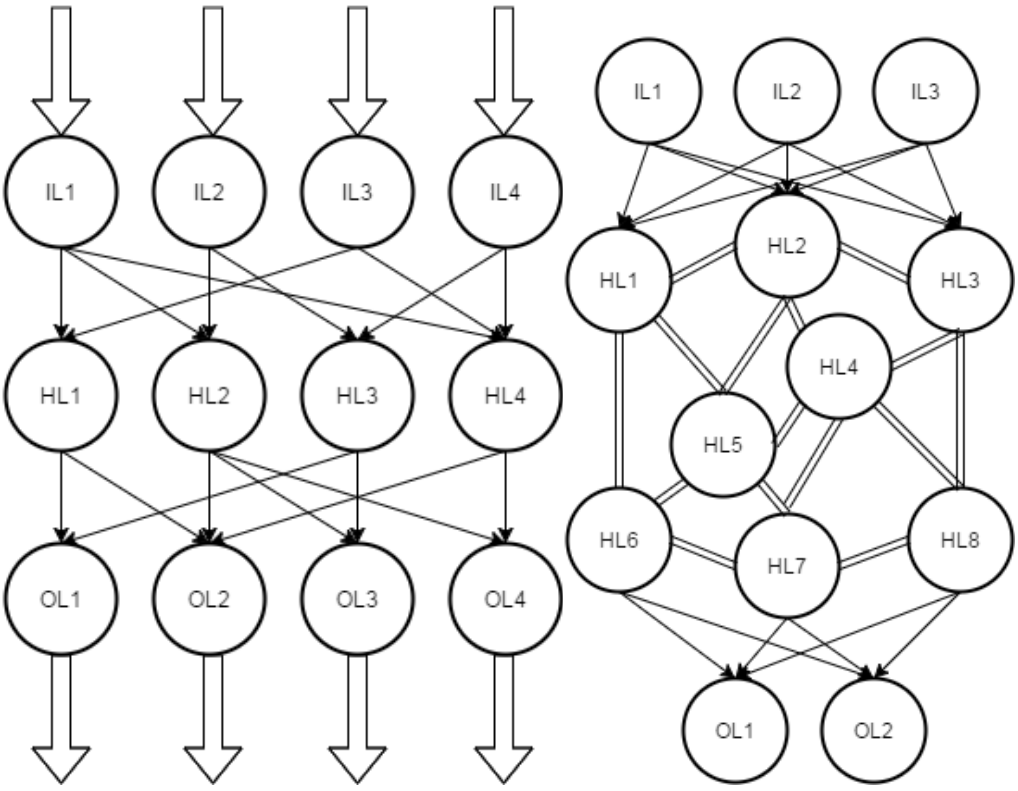
$$(\mathbf{J}^T\mathbf{J} + \lambda \mathbf{diag}(\mathbf{J}^T\mathbf{J}))\delta = \mathbf{J}^T[y - f(\mathbf{x})] \tag{2.20}$$

where λ is the damping factor and δ is the increment. An example of the Levenberg-Marquardt algorithm applied to the $1/f^{\beta}$ process is given in Figure 2.7. For all three β values that generate a different colored noise, it is possible to see that the algorithm fits the curve perfectly. This algorithm is the one chosen for the neural network model training generation.

2.2.2.2 Neural Networks .

The basic biological definition of a neural network is very similar to the definition of a neural network in machine learning. Regarding the latter, a neural network is a computer system modeled on the human brain and nervous system of the human body. It allows the studied system to learn from mistakes made and be allowed to improve the results through what is called the epochs [85–87] . There exist several neural networks, the ones chosen in this study were the feedforward neural network and the layer recurrent neural network.

Feedforward Neural Network - The simplest of all artificial neural networks, the feedforward neural network has only one layer and contrarily to the recurrent version, it does not cycle. The information only moves forward from the input nodes to the output nodes



(a) Feedforward Neural Network (b) Layer Recurrent Neural Network

Figure 2.8: Examples of feedforward and layer recurrent neural networks

through the hidden layer and it is considered to be a universal approximator [85, 87, 88]. Its use in this study is mainly for comparative purposes in order to visualize the effectiveness of a basic neural network to a more complex and well suited version such as the layer recurrent neural network introduced below. In this study, the number of nodes chosen are 150 as it seems that this outputs less errors than for any other number of nodes. An example of such a network can be found in Figure 2.8a.

Layer Recurrent Neural Network - As opposed to the neural network presented hereinabove, the layer recurrent neural network does actually form a cycle, allowing for a usage of their internal memory to process arbitrary sequences of inputs [86, 88–92]. Recurrent neural networks are considered to be the optimal choice for time series prediction thanks to the cycle that they exhibit [88, 91, 92]. In this study, the hidden sizes are chosen to be equal to 15. These are the default values in Matlab. An example of such a network can be found in Figure 2.8b.

2.2.2.3 Decision Trees .

Trees are structures used to predict the response to inputs. This method is often used in machine learning theory [79,93,94] and it is based on the construction of a binary tree where the nodes represent the tests made on the inputs. In this study, the regression tree and the bagged decision tree have been chosen as the structures used for prediction. They both are very useful since they provide easy to understand predictions in all types of conditions [93–95].

Regression Tree - Regression tree analysis is done when the predicted outcome can be considered a real number [79,93,94]. Prediction trees use the tree to represent the recursive partition. Each of the terminal nodes, or leaves, of the tree represents a cell of the partition, and has attached to it a simple model which applies in that cell only. The input is given in what is called the root node of the tree and then a sequence of questions is asked about the features. The interior nodes are labeled with questions, and the edges or branches between them labeled by the answers. The questions asked depend highly on the previous answers [79,93,94]. There are a few advantages for choosing a regression tree:

1. Predictions are hastily computed.
2. Understanding which variables have a considerable impact on making the decisions is rendered easy.
3. If data is missing, there might be a chance that the end of the tree is not reached. Fortunately, a prediction can still be made by averaging all the leaves in the sub tree.
4. There are fast, reliable algorithms to learn these trees such as the Levenberg-Marquardt algorithm.

Bagged Regression Tree - Orignally called bootstrap aggregated trees, bagged regression trees are designed to improve the stability and accuracy of the regular regression tree. First designed by Breiman in 1994, it leads to improvement for unstable procedures but can degrade the performance of more stable methods [95]. Definition 2.2.2 gives a general description of bagging [95].

Definition 2.2.2. *Bagging* Given a set of data S of size N, bagging generates M new training sets S_i each of size N_1 by sampling from S uniformly and with replacement. By sampling with replacement, some observations may be repeated in each S_i. This kind of sample is called bootstrap sample. The M models are fitted using the above M bootstrap samples and combined by averaging the output for regression models.

In this study, the bagged regression tree was set to having a minimum number of leaves at a low number.

2.2.3 AR Model

Forecasting is a process that involves the prediction of a signal or process by creating points outside of the bounds of the signal. Many methods exist that allows forecasting, some of them already mentioned previously such as the neural networks and the decision trees. Another method that was left unmentioned is the autoregressive model, or AR model. It is a representation of a type of random process which implies that it is greatly suited for time series manipulation. The AR model is considered to be a form of stochastic difference equation. If p is the order of the AR model, the latter can defined as [96]:

$$X_t = cst + \sum_{n=1}^{p} \phi_n X_{t-n} + \epsilon_t \tag{2.21}$$

where ϕ_n, $\forall n \in [\![1, n]\!]$, are the parameters of the AR model and ϵ_t is white noise. Once the parameters X_t have been estimated, the AR model can be used to forecast an arbitrary number of periods into the future. There are four sources of uncertainty when using the AR model as a predictor [96, 97]:

1. Is the AR model the the correct model ?
2. Is the accuracy of the forecasted values used as lagged values in the AR equation good enough ?
3. Are the values of the AR coefficients good enough ?
4. Is the value of ϵ_t for the period being predicted correct ?

Points 2 to 4 are quantified and used to give what is called a confidence interval which widens the more points are being forecasted. This is due to the increasing number of estimates the algorithm lies on to make the predictions [96, 97].

There is therefore doubt on the actual accuracy and quality of the forecasts using the AR model. However, for low parameter values it can be considered a good enough method for prediction, hence its use in this study. Its main purpose will be mostly to compare it to the machine learning algorithms as a way to evaluate the accuracy and functionality of these aforementioned methods.

3.0 METHODOLOGY

3.1 EVALUATING $1/f^{\beta}$ PROCESS

3.1.1 Stride Interval Definition

Since the $1/f^{\beta}$ power law recreates the stride interval time series, the best way to get results that recreate real life conditions is to simulate the power law [61, 98]. In this study, the number of strides chosen was 300 as it is the average number of strides a healthy individual does in five minutes that is not considered to be active in sports [4, 5, 18]. As a reference, an athlete does 190 steps per minute. Obviously a person that is affected by a disease that declines his locomotor system or an older adult will be able to achieve a lower amount of strides in that same amount of time and the spacing between the strides is much larger. However, since the aim of this study is to be able to predict the $1/f^{\beta}$ process, these constraints were disregarded. The method on how to reproduce that process is discussed in the following subsection.

3.1.2 Generating Discrete $1/f^{\beta}$ Process

First of all, since the $1/f^{\beta}$ process was needed, the generation of the latter was achieved. There exist some methods that allow to accurately simulate the process. In this discussion, a chronological presentation will be briefly done. A first method was proposed by Li and other of his associates where he created a filter of fractional order in order to generate fBm processes with stochastically fractional differential equations [37, 49, 51]. That method was extended and generalized for any filter that would include fGn and fBm signal, or $1/f^{\beta}$ processes [50]. This method was implemented for this numerical analysis of $1/f^{\beta}$ processes.

According to Kasdin, the transfer function of the fractional system that follows the power law of β is given by introducing a series $h(n)$ that can be calculated using an iterative process and convoluting the result to white Gaussian noise in order to obtain the realization of the process $x(n)$[50]. A mathematical demonstration is given below [50]:

$$h(n) = \frac{\Gamma(\beta/2 + n)}{n!\Gamma(\beta/2)} \tag{3.1}$$

which is calculable using the following iterative process, where $h(0) = 1$:

$$h(n) = \left(\frac{\beta}{2} + n - 1\right) \frac{h(n) - 1}{2} \tag{3.2}$$

Finally, the realization of the process $x(n)$ is given by:

$$x(n) = wgn(n) * h(n) \tag{3.3}$$

where $*$ denotes the convolution symbol and $wgn(n)$ is white Gaussian noise.

3.2 ANALYSIS SCHEME

3.2.1 Analysis of Simulated Time-Series: Monte Carlo simulation

The basis of this computational evaluation is the generation of $1/f^{\beta}$ processes. For completeness, the prediction models and forecasted points were calculated using the smaller range of β, i.e. $0 \leq \beta \leq 2$. Indeed, due to physiological constraints, the anti-persistent fractional Brownian motion, $\beta > 2$ and the anti-correlated fractional Gaussian noise, $\beta < 0$, were disregarded. Thus, this only includes a smaller range of possible values of the Hurst exponent, $0 < H < 1$, meaning that it still includes the fractional Gaussian noise and the fractional Brownian motion but not the unwanted possibilities. As previously mentioned, the methods are evaluated for a total number of points equal to 300 strides in order to recreate healthy conditions as a basis. Moreover, it is also calculated over a range of β values and since predictions can sometimes be inaccurate after a certain number of points, usually very low, the number of points forecasted include 5 points, then 15, 25, 35, 50, 100 and finally

150 points forecasted. More precisely, for forecasting an amount N of strides, or points, the prediction would start at $300 - N$ strides in order to achieve a total of 300 strides. Given the stochastic nature of these processes, it is important determine a convergent value on which each technique forecasts the time-series. The procedure of signal generation and calculation is implemented in a Monte Carlo scheme, where each realization is repeated 100 times. Since it is wished to verify the accuracy of the predictions made, every calculations are made for a starting β value at 0 following and incrementation of 0.1 at each iteration, i.e. $\beta = 0, 0.1, 0.2, ..., 2$. In each iteration for a set of β values, the time-series is normalized and evaluated by each of the machine learning methods following extrapolation and then a mean of these 100 realizations is computed in order to get a single signal representative of these signals for a specific β value. Importantly, each iteration is carried out for the same realization of randomly generated $1/f^{\beta}$ process.

3.2.2 Evaluating the Processes

The best way to evaluate the accuracy of the modeled $1/f^{\beta}$ processes after the machine learning algorithms is to calculate the mean and the coefficient of variation of these models. These methods were implemented in the Monte Carlo simulation mentioned hereinabove. The mean allows to observe the long-run expected average. If the models after simulation have a mean that can be considered roughly similar to the mean of the actual three hundred stride signal then it can be said that the signal that exhibits this feature is a good enough model for the number of forecasted points it predicts. Stride interval variability (SIV) was estimated using the coefficient of variation of stride intervals [99, 100]. The latter is simply a measure of dispersion of a given variable or process, it shows the extent of variability in relation to the mean of the population. Both of these linear measures characterize the amount of variability present in the data [101]. However, there is an advantage in using the coefficient of variation, it can compare across different variables or processes because they are now measured on the same relative scale (ratio). Furthermore, the standard deviation of data must always be understood in the context of the mean of the data. A process with the smaller coefficient of variation is less dispersed than the process with the larger coefficient

of variation. Generally, changes in the coefficient of variation are indicative of increases or decreases in the amount of variability [101]. However, if the mean of a signal is close to zero, the value of the coefficient of variation will approach infinity. Therefore any small changes in the mean of a signal will significantly impact the value of the coefficient of variation. It is a tool that is used in many fields, namely in statiscal and applied probability and its definition is simple. For a signal x, the coefficient of variation is defined as follows [101]:

$$CoV(x) = 100\frac{\sigma(x)}{\mu(x)} \tag{3.4}$$

where $\sigma(x)$ denotes the standard deviation of the signal x and $\mu(x)$ is its mean. The ratio is multiplied by 100 in order to get a percentage reference of the coefficient of variation. The calculation of the coefficient of variation and the mean in the Monte Carlo simulation is proven to be deterministic when it comes to time series prediction [100, 101]. If a predicted signal presents similar values as the actual signal, then the former is an acceptable prediction model.

4.0 RESULTS

In this section are presented the error plots of the results of the simulation using the time series compared to the original signal. From these results this thesis seeks to determine which of the several method presented hereinabove can most effectively and accurately evaluate the prediction of the simulated time series under various constraints. a great number of points forecasted being the ultimate goal, the aim of this study is also determine up to which number of points is it possible to get an accurate measure. The stride interval time series was simulated using the fact that they can be simulated using the the $1/f^{\beta}$ power law. As mentioned, the number of strides is 300 and the tests were conducted using $0 \leq \beta \leq 2$ in order to mimic a spectrum of conditions. A Monte Carlo simulation was computed where a forecast of 5, 15, 25, 35, 50, 100 and 150 points was done compared to the spectral exponent. The results are presented in the two following sections. The mean of the simulated and modeled time series will firstly be compared followed by the one for the coefficient of variation.

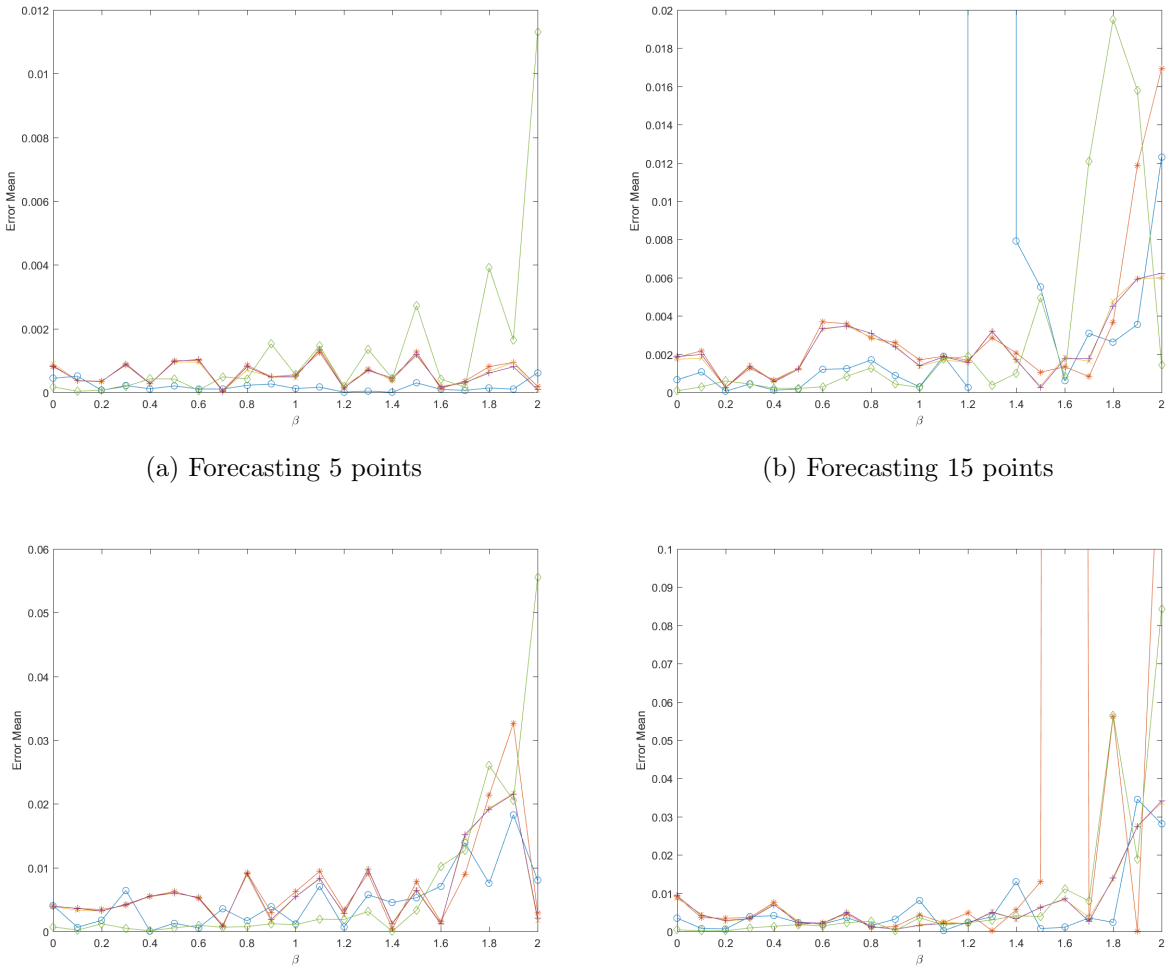
4.1 COMPARISON OF THE MEAN

The results in this section were acquired through a Monte Carlo simulation using the conditions mentioned in Chapter 3. The $\beta \in [0, 2]$ spectrum presents the performance of the general scheme which allows to calculate the mean of the entirety of the forecasted signals with significance. Shown are the results of the simulated time series along with the forecasted signals for a variety of total forecasted points over 300 strides in Figures 4.1 and 4.2. Figure 4.1a corresponds to 5 points being forecasted, as for Figures 4.1b, 4.1c, 4.1d, 4.2a, 4.2b and 4.2c, they respectively correspond to 15, 25, 35, 50, 100 and 150 points being forecasted.

The noticeable trait on all signals is that regardless of the conditions, the mean of simulated forecasts correspond with the mean of the original signal, although some do have noticeable effects as discussed below. The exception is for the layer recurrent neural network. Although it does follow the same pattern as the others, the mean varies greatly compared to the others. By looking into more details, it can be seen that the error between the mean of this model and that of the original signal is fairly low when the spectral exponent β is lesser than 1. As the value of β increases though, the mean value changes as well. The difference increases drastically where in some cases the peaks reach an immeasurable value as seen in some figures where some values are outside the bounds, however for the LRNN, the maximum value reached is well above that of the original signal as seen in Figure 4.2c for example. The number of points forecasted does exhibit some impact in the mean values but the general pattern is still the same, apart for the LRNN. Some peaks are visible in Figures 4.1b for FFNN and 4.1d for the LRNN. It can be the consequence of calculation errors since it only happens for one particular β value in one set of forecasted points. These changes, albeit quite benign, are under the form of small increases in the mean value for some method, most notably for the layer recurrent neural network.

In order to get a better view of the results exhibited in Figure 4.1, numerical values can be seen in Table 4.1. It displays the values (*mean* $\pm$ *standard deviation*) of the difference between a modeled and the actual signal for a given number of points being forecasted. It can be seen that the largest values are displayed for the largest number of points forecasted. Furthermore, this further shows that the methods used does have an impact on the mean value where some are greater than others. The most interesting values are obtained with both the regression tree and bagged regression tree, and the auto-regressive method. The feedforward neural network does display interesting results as well however due to the small error peaks that often appears when forecasting does not make it reliable enough for predictions. The results in the table regarding the latter is conclusive, the feedforward neural network exhibits greater differences with the original signal compared to the AR model and both decision trees.

A second series of test was done in order to have more conclusive results. The following subsection will display the results obtained by using the coefficient of variation. A comparison will then be made between the results using both methods.



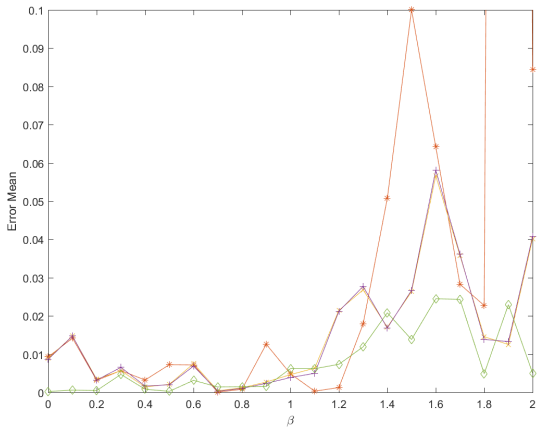
(a) Forecasting 5 points

(b) Forecasting 15 points

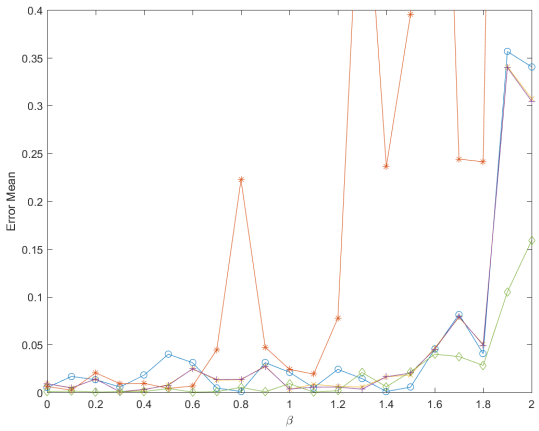
(c) Forecasting 25 points

(d) Forecasting 35 points

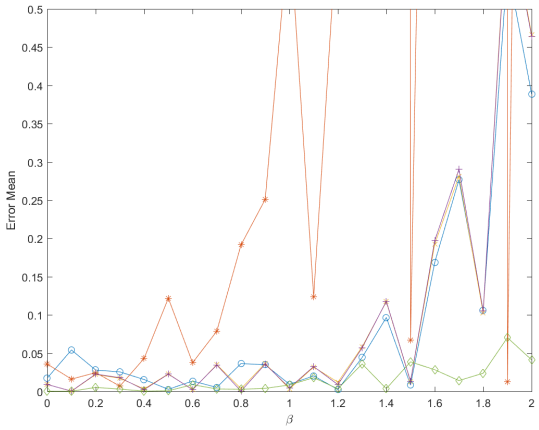
Figure 4.1: Part 1: Error plot of the MonteCarlo simulation result for the mean of the forecasted signals depending on the β value compared to the original signal. Signals: -o is FFNN, -* is LRNN, -x is the regression tree, -+ is the bagged regression tree, $\Diamond$ is the AR model



(a) Forecasting 50 points



(b) Forecasting 100 points



(c) Forecasting 150 points

Figure 4.2: Part 2: Error plot of the MonteCarlo simulation result for the mean of the forecasted signals depending on the β value compared to the original signal. Signals: -o is FFNN, -* is LRNN, -x is the regression tree, -+ is the bagged regression tree, $\Diamond$ is the AR model

Table 4.1: Absolute value of the difference of the mean of the models and the actual signal when $0 \leq \beta \leq 2$. FFNN: Feedforward Neural Network, LRNN: Layer Recurrent Neural Network, RT: Regression Tree, BRT: Bagged Regression Tree, AR: Autoregressive Model

	FFNN	**LRNN**	**RT**	**BRT**	**AR**
5 points ($\times 10^{-5}$)	2.5452±24.952	3.7428±72.836	0.6202±71.107	3.8314±70.646	111.29±259.91
15 points	1.5×10^{6} $\pm 6.9\times10^{6}$	0.0217±0.1474	0.0227±0.1485	0.0226±0.1485	0.0265±0.1532
25 points	0.0189±0.1301	0.0192±0.1339	0.0192±0.1320	0.0192±0.1319	0.0229±0.1306
35 points	0.0236±0.1276	0.1394±0.5228	0.0218±0.1299	0.0218±0.1299	0.0309±0.1248
50 points	18.540±9.5076	0.0390±0.2545	0.0050±0.0207	0.0237±0.0573	0.0206±0.0575
100 points	0.0489±0.1806	0.0410±0.5812	0.0486±0.1724	0.0484±0.1718	0.0347±0.1137
150 points	0.0834±0.2563	362.12±1515.3	0.0898±0.2757	0.0902±0.2747	0.0145±0.1556

4.2 COMPARISON OF THE COEFFICIENT OF VARIATION

In order to strengthen what is already know, the coefficient of variation was calculated for all modeled signals along with the forecast as well as the original signal. The results in this section were also acquired through a Monte Carlo simulation where the same conditions were applied. Shown are the error plots of the results of the simulated time series compared to the forecasted signals for a variety of total forecasted points over 300 strides in Figures 4.3 and 4.4. Figure 4.3acorresponds to 5 points being forecasted, as for Figures 4.3b, 4.3c, 4.3d, 4.4a, 4.4b and 4.4c, they respectively correspond to 15, 25, 35, 50, 100 and 150 points being forecasted.

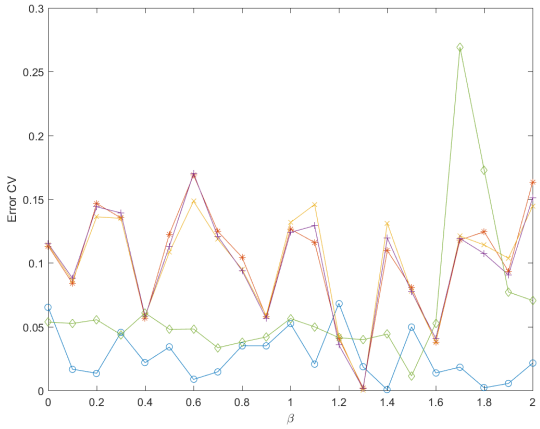
It can be seen that the value of the spectral exponent β does have a huge impact on the mean and coefficient of variation. As the value of β increases, the error values of the coefficient of variation of the original signal and that of the modeled signal increase as well. The accuracy regarding the coefficient of variation is very different than the one for the mean since the error here is far greater. That error oscillates near 0 for all methods but as soon as β verifies $\beta > 1$, although that value changes depending on the number of points being forecasted, the error increases drastically, reaching nearly 35 when forecast a greater number of strides which is far greater than the errors when forecasting smaller numbers. However, as can be seen in both Figures 4.3 and 4.4, that value is not the maximum that is reached since there are some signals that have far greater maxima than that of the original signal. Since these signals' range are so great, a consideration was made where the acceptable bound would be maximum thirty. Outside these bounds, the error is considered too great to be considered accurate and thus the value at the corresponding β value is disregarded.

A small discrepancy in the coefficient of variation values for each method is visible when $\beta \leq 1$ but the gap only widens when the number of forecasted points increases. If a more separate case is done, regarding the coefficient of variation results, one can see that all methods display a signal that is similar to the original, with very few differences when the number of points increases. Above 25 forecasted points, it becomes apparent the layer recurrent neural network starts to be unreliable as its coefficient of variation values dramatically increase. A

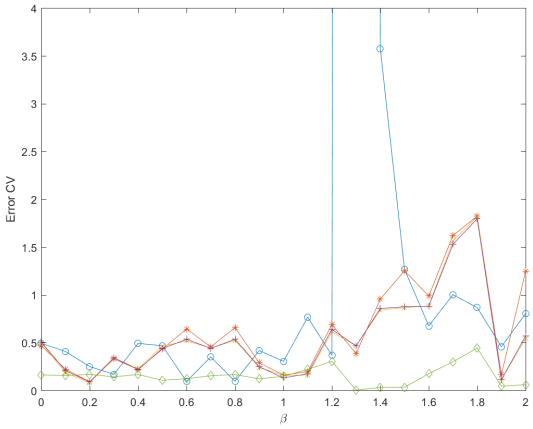
noteworthy information is that increase starts when the signal is no longer considered pink noise, but rather starts to resemble a fractal Brownian motion type signal, or red noise. In Fig 4.4c, the coefficient of variation of the LRNN is entirely different from the original signal, yet it still oscillates around the values of it. Similarly to the results found in Figure 4.1b, the FFNN in Figure 4.3b displays a peak at the same β value as with the mean. This can be explained by the definition of the coefficient of variation mentioned previously.

The FFNN, regression tree, bagged regression tree, and the AR have considerably better approximations than the LRNN when forecasting over 25 points. However, when exceeding 50 points, the predicted signals, albeit relatively similar, have coefficients of variation that are greater than that of the original, albeit relatively similar. Only the AR model has a far more accurate prediction than the others as seen in Figure 4.4c.

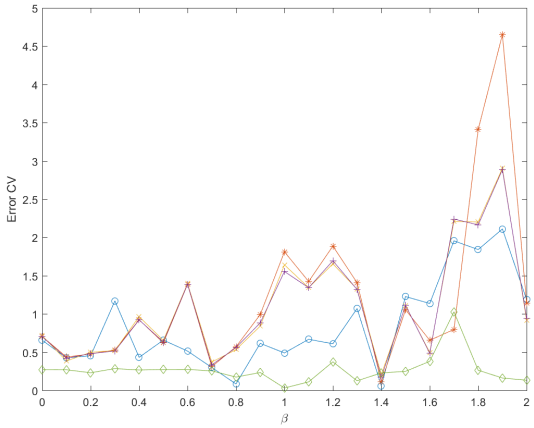
Similarly to the previous section, a table was designed in order to get a general view of the results shown in Figure 4.3. These numerical values can be found in Table 4.2. It displays the values (*mean* $\pm$ *standard deviation*) of the difference between a modeled and the actual signal for a given number of points being forecasted. It can be seen that the layer recurrent neural network does not display close enough values to the actual signal as seen in the figures. It has an average coefficient of variation that well above what is expected for the predictions. Similarly, the bagged regression tree, regression tree and layer recurrent neural network have a slightly greater coefficient of variation but does not come close enough to the actual signal compared to the autoregressive model. Regarding machine learning methods, the decision trees can be considered better suited for the task, however the AR model would generally be the best method since it is not affected by the number of points forecasted and has the closest values to the ones of the actual signal.



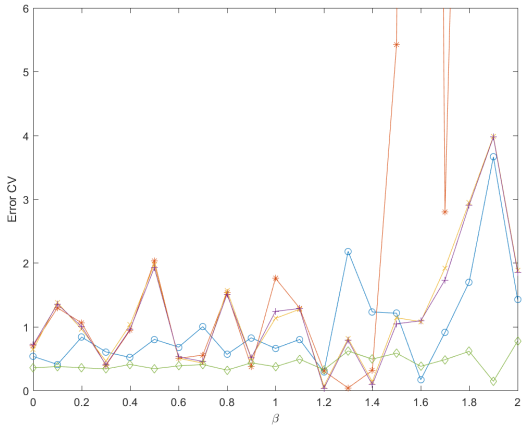
(a) Forecasting 5 points



(b) Forecasting 15 points

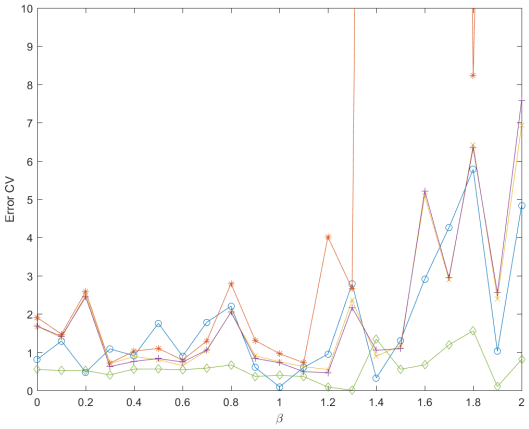


(c) Forecasting 25 points

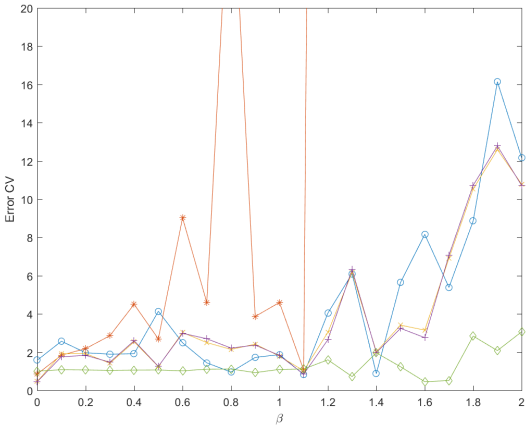


(d) Forecasting 35 points

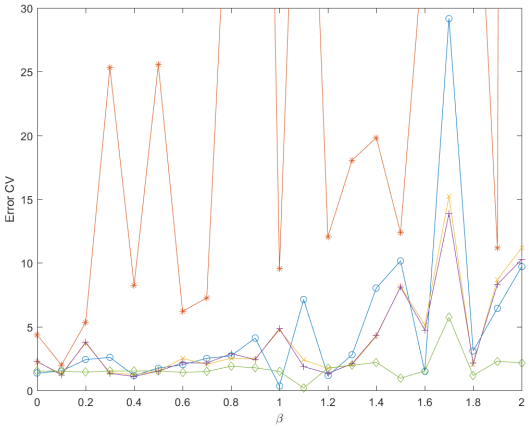
Figure 4.3: Part 1: Error plot of the MonteCarlo simulation result for the coefficient of variation of the forecasted signals depending on the β value compared to the original signal. Signals: -o is FFNN, -* is LRNN, -x is the regression tree, -+ is the bagged regression tree, $\Diamond$ is the AR model



(a) Forecasting 50 points



(b) Forecasting 100 points



(c) Forecasting 150 points

Figure 4.4: Part 2: Error plot of the MonteCarlo simulation result for the coefficient of variation of the forecasted signals depending on the β value compared to the original signal. Signals: -o is FFNN, -* is LRNN, -x is the regression tree, -+ is the bagged regression tree, $\Diamond$ is the AR model

Table 4.2: Absolute value of the difference of the coefficient of variation of the models and the actual signal when $0 \leq \beta \leq 2$. FFNN: Feedforward Neural Network, LRNN: Layer Recurrent Neural Network, RT: Regression Tree, BRT: Bagged Regression Tree, AR: Autoregressive Model

	FFNN	**LRNN**	**RT**	**BRT**	**AR**
5 points	0.0228±0.0243	0.0648±0.0905	0.0651±0.0884	0.0653±0.0881	0.0059±0.0865
15 points	6.9217±31.862	0.0207±3.1158	0.0574±3.2040	0.0553±3.2022	0.6975±3.1601
25 points	0.2798±3.7759	0.6388±3.6827	0.5141±3.8207	0.5114±3.8228	0.6160±3.9940
35 points	0.5215±1.8920	4.9118±12.221	0.7452±1.9041	0.7273±1.9122	0.7917±2.1511
50 points	1.7148±1.5854	8.9867±100.61	2.0304±1.8839	2.0497±1.9855	0.3006±0.6439
100 points	4.5375±0.2016	23.669±211.57	4.0746±3.7515	4.0519±3.7912	0.0777±2.0105
150 points	4.9999±6.9583	12.529±137.55	4.2190±4.4150	4.0803±4.1100	0.6692±3.2485

5.0 DISCUSSION

In this study, we considered five methods, which includes four machine learning methods, to predict the stochastic process that is the $1/f^{\beta}$ function. A Monte Carlo simulation was done for all five methods on a range of β values for a given number of forecasted points ranging from 5 to 150. The mean of the signals as well as their coefficient of variation have been computed. Both were used for the simulation. In this section, an in-depth approach will be conducted in order to better understand the results that were obtained.

5.1 PREDICTED SIMULATED MODELS

The results regarding the mean of the signals obtained through the simulation have shown that the regression tree, bagged regression tree, feedforward neural network and the autoregressive model are well suited for predicting the $1/f^{\beta}$ process, however the FFNN was deemed unreliable due to the errors it presented. As for the simulation using the coefficient of variation, the same signals also exhibited promising predictions. By comparing these results, the FFNN can no longer be considered a viable option. From the remaining prediction methods, only the decision trees and the AR model can be considered for forecasting purposes.

Since the $1/f^{\beta}$ process is a stochastic process, it was not surprising that predicting it would not give accurate results. When using the neural networks and the decision trees, the predictions do not display the exact variations of the white, pink and red noises. This resulted

in choosing the regression trees as the best methods for predicting the stride interval time series, as well as the AR model. However, when forecasting less than 25 points, all methods used in this study can be deemed acceptable for this purpose, yet the most robust are still the ones mentioned previously. Since the goal is to predict the future β value of an individual using the Average Wavelet Coefficient (AWC), either the RT, BRT or the AR model suffice. However, as mentioned previously, if the goal is to get the value one particular point at a future state, the best method would therefore be the AR model since it is the most robust of all the methods presented in this study.

5.2 INFLUENCE OF THE BIOMARKER

The biomarker β plays an essential role in the prediction outcome. As seen in the Monte Carlo simulations, its value greatly influences the accuracy of the models. If the value is such that $1 < \beta \leq 2$, the signals used for prediction might be unreliable, depending on the number of points that is being forecasted. Previous studies, most notably with Hausdorff et al., have shown that these β values are more often than not inferior to one, which is considered as being the benchmark for a healthy patient, regardless of the health state of the individual (affected by Huntington's disease, Parkinson's disease or ALS in the case of the study) [5,17,18]. However, this piece of information does not alter the conclusions drawn previously. It is in that range of β values, $0 \leq \beta \leq 1$, that the simulated signals exhibit more accuracy.

In this study, the prediction was only performed on simulated data which indeed helps give an idea of the outcome, however data obtained through acquisition on real life cases helps give more probable results. The simulated data does not take into account the difference in stride intervals between individuals, it is constant regardless of the β value. Furthermore, the Monte Carlo simulation was only done with 100 simulation per β value. Although it can give accurate results, an increased number of simulations would only strengthen the conclusions that are being drawn. Finally, only a handful of methods were used in for predicting the

signals. Machine learning methods are well suited for prediction and forecasting, there exist many different types of neural networks and decision trees, the ones chosen in this study are the more common ones.

In order to improve the results obtained in this study, it would be interesting to apply what is already known here to real life data, as well as increase the number of simulations to 1000. That would require a lot of computational power since the simulation is voracious time-wise and computer-wise with only 100 simulations per β value. Finally, it would be interesting to explore more machine learning methods for the purpose of increasing the total number of points that is possible to forecast.

6.0 CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

The object of this study was to provide a way to predict the non-zero mean $1/f^{\beta}$ time series with respect to the physiological applications as well as determining a number of points that can be forecasted accurately in an acceptable way. Primarily, a prediction analysis was carried out in order to determine the functionality of various prediction methods available. If fractal characteristics are of any interest for some arbitrary physiological process, it is important to determine which method may output acceptable accurate results as this can prove useful in the medical field. Predicting stochastic processes proves to be quite difficult which is why the choice of prediction methods is crucial. Machine learning and regression methods prove to do well when dealing with such processes as this study has shown. Two neural networks were studied, the feed forward neural network and the layer recurrent neural network, as well as two regression trees, the regression tree and the bagged regression tree, and finally the autoregressive (AR) model. The evaluation of the neural networks have shown that for a high β value and a high number of strides forecasted, they cannot be applied for extrapolation. However, these are still relatively valid evaluations if these neural networks are applied in the special conditions where the spectral exponent is between 0 and 1 and that the number of strides predicted is below a certain value. On the other hand, both regression trees and the AR model present a far better and accurate result for forecasting points. The previously carried out evaluations have shown that a more accurate result is given for an total number of forecasted points around 15 strides which is satisfactory given the methods at hand, however it still remains a low value compared to the three hundred initial strides. This study is the first of its kind that applies the time series prediction on stride interval time

series thus comparisons to other studies cannot be made. However, this paves the way for other studies to be made in the field of stride interval time series forecasting by introducing the importance of such studies and how it can affect the medical world thanks to the notions and results they carry.

6.2 FUTURE WORK

This investigation uncovered some of the pitfalls of some methods of prediction in fractal time series analysis. Any recommendations of algorithm modification or development were out of the scope of this comparative and forecasting assessment. For example, there are many ways in which neural networks and regression trees, or other machine learning and regression algorithms, may be improved to more suitably predict and forecast the modeled non-zero mean $1/f^{\beta}$ stride interval time series process. In the interest of developing the general scheme, parameters such as the total number of stride intervals, the spacing between strides or the number of points forecasted, were kept constant and did not necessarily apply to all the possible cases. As discussed previously no other studies have been conducted on the gait stride interval time series prediction, it would thus be interesting to extend and refine the analysis done in this thesis in order to forecast even more points accurately so that specific mechanisms of disease and decline in gait can be better observed and understood. Also, this study focused on a simplified assumption of the complex phenomena of interest in that all processes are assumed to be monofractal which is presently unclear if they actually are. Although some multifractal analysis in physiology have been performed, applying it to the stride interval time series would prove to be interesting. Furthermore, a more widespread use of fractal measures as a parameter in other investigations of nonlinear and nonstationary phenomena would be interesting to study. It can prove to be a useful tool in studies that focus on pattern recognition problems in dysphagia or for EEG measures. Finally, it seems that model predictions would be well suited to inform the development of more behaviorally accurate models of complex physiological systems. With technology always improving and implemented in our day to day life, a more mobile version for verifying the physiological

state of one individual that can efficiently output results would prove to be useful. It would be interesting to investigate the development of one such mobile device and implement it in rehabilitation centers where patients with a pathological gait due to accidents, for example, learn to walk again.

7.0 ACKNOWLEDGEMENT

I would first like to thank my parents Patrick and Brigitte for being there for me even though they live in a different country. It was not easy but their love and support greatly helped in every way shape or form. My thoughts also go to my Godfather Joël Le Gall for having the generosity to welcome me here and essentially take care of me during my studies in Pittsburgh. Outside my immediate family, I would like to thank everyone who helped me and supported during this year outside my country. This opportunity was given to me thanks to my French engineering ENSEA and Dr. El-Nokali of the University of Pittsburgh Swanson School of Engineering. I would like to also thank all my friends at the lab for offering me a lot of help in this research thanks to their knowledge of the subject and how to solve technical issues. Finally, I would like to thank my advisor Dr. Sejdić for his patience, guidance and belief. His dedication to forming me is truly appreciated.

BIBLIOGRAPHY

[1] F. B. Horak, "Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls?" *Age and ageing*, vol. 35, no. suppl 2, pp. ii7–ii11, 2006.

[2] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Sensors*, vol. 12, no. 2, pp. 2255–2283, 2012.

[3] V. T. Inman, "Human locomotion," *Canadian Medical Association Journal*, vol. 94, no. 20, p. 1047, 1966.

[4] J. M. Hausdorff, C. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Is walking a random walk? evidence for long-range correlations in stride interval of human gait," *Journal of Applied Physiology*, vol. 78, no. 1, pp. 349–358, 1995.

[5] J. M. Hausdorff, P. L. Purdon, C. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations," *Journal of Applied Physiology*, vol. 80, no. 5, pp. 1448–1457, 1996.

[6] A. Eke, P. Herman, J. Bassingthwaighte, G. Raymond, D. Percival, M. Cannon, I. Balla, and C. Ikrényi, "Physiological time series: distinguishing fractal noises from motions," *Pflügers Archiv*, vol. 439, no. 4, pp. 403–415, 2000.

[7] A. Eke, P. Herman, L. Kocsis, and L. Kozak, "Fractal characterization of complexity in temporal physiological signals," *Physiological Measurement*, vol. 23, no. 1, p. R1, 2002.

[8] L. Glass, "Synchronization and rhythmic processes in physiology," *Nature*, vol. 410, no. 6825, pp. 277–284, 2001.

[9] R. W. Glenny, H. T. Robertson, S. Yamashiro, and J. B. Bassingthwaighte, "Applications of fractal analysis to physiology," *Journal of Applied Physiology*, vol. 70, no. 6, pp. 2351–2367, 1991.

[10] A. L. Goldberger and B. J. West, "Fractals in physiology and medicine." *The Yale Journal of Biology and Medicine*, vol. 60, no. 5, p. 421, 1987.

[11] J. P. Sturmberg and B. J. West, "Fractals in physiology and medicine," in *Handbook of Systems and Complexity in Health*. Springer, 2013, pp. 171–192.

[12] H. V. Huikuri, T. H. Mäkikallio, K. J. Airaksinen, T. Seppänen, P. Puukka, I. J. Räihä, and L. B. Sourander, "Power-law relationship of heart rate variability as a predictor of mortality in the elderly," *Circulation*, vol. 97, no. 20, pp. 2031–2036, 1998.

[13] H. V. Huikuri, T. H. Mäkikallio, C.-K. Peng, A. L. Goldberger, U. Hintze, M. Møller, and D. S. Group, "Fractal correlation properties of RR interval dynamics and mortality in patients with depressed left ventricular function after an acute myocardial infarction," *Circulation*, vol. 101, no. 1, pp. 47–53, 2000.

[14] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, no. 6735, pp. 461–465, 1999.

[15] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 5, no. 1, pp. 82–87, 1995.

[16] J. Hausdorff, L. Zemany, C.-K. Peng, and A. Goldberger, "Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children," *Journal of Applied Physiology*, vol. 86, no. 3, pp. 1040–1047, 1999.

[17] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *Journal of Applied Physiology*, vol. 88, no. 6, pp. 2045–2053, 2000.

[18] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C.-K. Peng, M. E. Cudkowicz, J. Y. Wei, and A. L. Goldberger, "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and huntingtons disease," *Journal of Applied Physiology*, vol. 82, no. 1, pp. 262–269, 1997.

[19] K. Sharma, J. Kent-Braun, S. Majumdar, Y. Huang, M. Mynhier, M. Weiner, and R. Miller, "Physiology of fatigue in amyotrophic lateral sclerosis," *Neurology*, vol. 45, no. 4, pp. 733–740, 1995.

[20] K. R. Sharma and R. G. Miller, "Electrical and mechanical properties of skeletal muscle underlying increased fatigue in patients with amyotrophic lateral sclerosis," *Muscle and Nerve*, vol. 19, no. 11, pp. 1391–1400, 1996.

[21] N. Scafetta, D. Marchi, and B. J. West, "Understanding the complexity of human gait dynamics," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 19, no. 2, p. 026108, 2009.

[22] O. Blin, A.-M. Ferrandez, and G. Serratrice, "Quantitative analysis of gait in parkinson patients: increased variability of stride length," *Journal of the Neurological Sciences*, vol. 98, no. 1, pp. 91–97, 1990.

[23] J. B. Bassingthwaighte, "Physiological heterogeneity: fractals link determinism and randomness in structures and functions," *Physiology*, vol. 3, no. 1, pp. 5–10, 1988.

[24] J. B. Bassingthwaighte and R. P. Bever, "Fractal correlation in heterogeneous systems," *Physica D: Nonlinear Phenomena*, vol. 53, no. 1, pp. 71–84, 1991.

[25] T. Chau, "A review of analytical techniques for gait data. part 1: Fuzzy, statistical and fractal methods," *Gait and Posture*, vol. 13, no. 1, pp. 49–66, 2001.

[26] D. Delignières and K. Torre, "Fractal dynamics of human gait: a reassessment of the 1996 data of Hausdorff et al." *Journal of Applied Physiology*, vol. 106, no. 4, pp. 1272–1279, 2009.

[27] S. M. Bruijn, D. J. Bregman, O. G. Meijer, P. J. Beek, and J. H. van Dieën, "Maximum lyapunov exponents as predictors of global gait stability: a modelling approach," *Medical Engineering and Physics*, vol. 34, no. 4, pp. 428–436, 2012.

[28] D. Delignières, M. Fortes, and G. Ninot, "The fractal dynamics of self-esteem and physical self," *Nonlinear Dynamics in Psychology and Life Sciences*, vol. 8, pp. 479–510, 2004.

[29] D. Delignieres, S. Ramdani, L. Lemoine, K. Torre, M. Fortes, and G. Ninot, "Fractal analyses for shorttime series: a re-assessment of classical methods," *Journal of Mathematical Psychology*, vol. 50, no. 6, pp. 525–544, 2006.

[30] S. Havlin, S. Buldyrev, A. Goldberger, R. Mantegna, S. Ossadnik, C.-K. Peng, M. Simons, and H. Stanley, "Fractals in biology and medicine," *Chaos, Solitons and Fractals*, vol. 6, pp. 171–201, 1995.

[31] S. Bruijn, O. Meijer, P. Beek, and J. Van Dieën, "Assessing the stability of human locomotion: a review of current measures," *Journal of the Royal Society Interface*, vol. 10, no. 83, p. 20120999, 2013.

[32] M. M. Mielke, R. O. Roberts, R. Savica, R. Cha, D. I. Drubach, T. Christianson, V. S. Pankratz, Y. E. Geda, M. M. Machulda, and R. J. Ivnik, "Assessing the temporal relationship between cognition and gait: slow gait predicts cognitive decline in the mayo clinic study of aging," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 68, no. 8, pp. 929–937, 2013.

[33] H. H. Atkinson, C. Rosano, E. M. Simonsick, J. D. Williamson, C. Davis, W. T. Ambrosius, S. R. Rapp, M. Cesari, A. B. Newman, and T. B. Harris, "Cognitive function, gait speed decline, and comorbidities: the health, aging and body composition

study," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 62, no. 8, pp. 844–850, 2007.

[34] Y. A. Grimbergen, M. J. Knol, B. R. Bloem, B. P. Kremer, R. A. Roos, and M. Munneke, "Falls and gait disturbances in huntington's disease," *Movement Disorders*, vol. 23, no. 7, pp. 970–976, 2008.

[35] C. J. Lamoth, F. J. van Deudekom, J. P. van Campen, B. A. Appels, O. J. de Vries, and M. Pijnappels, "Gait stability and variability measures show effects of impaired cognition and dual tasking in frail people," *Journal of Neuroengineering and Rehabilitation*, vol. 8, no. 1, p. 1, 2011.

[36] R. Bryce and K. Sprague, "Revisiting detrended fluctuation analysis," *Scientific Reports*, vol. 2, 2012.

[37] M. Li, "Fractal time seriesa tutorial review," *Mathematical Problems in Engineering*, vol. 2010, 2009.

[38] B. B. Mandelbrot and J. W. Van Ness, "Fractional brownian motions, fractional noises and applications," *Journal of the Society for Industrial and Applied Mathematics*, vol. 10, no. 4, pp. 422–437, 1968.

[39] B. B. Mandelbrot, "Self-affine fractals and fractal dimension," *Physica Scripta*, vol. 32, no. 4, p. 257, 1985.

[40] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A: Statistical Mechanics and its Applications*, vol. 316, no. 1, pp. 87–114, 2002.

[41] C. Peng, S. Havlin, J. Hausdorff, J. Mietus, H. Stanley, and A. Goldberger, "Fractal mechanisms and heart rate control: Long-range correlations and their breakdown with disease," *Studies in Health Technology and Informatics*, pp. 3–14, 1997.

[42] M. F. Shlesinger, "Fractal time and $1/f$ noise in complex systems," *Annals of the New York Academy of Sciences*, vol. 504, no. 1, pp. 214–228, 1987.

[43] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality," *Physical Review A*, vol. 38, no. 1, p. 364, 1988.

[44] M. S. Taqqu, V. Teverovsky, and W. Willinger, "Estimators for long-range dependence: an empirical study," *Fractals*, vol. 3, no. 04, pp. 785–798, 1995.

[45] G. Rangarajan and M. Ding, "Integrated approach to the assessment of long range correlation in time series data," *Physical Review E*, vol. 61, no. 5, p. 4991, 2000.

[46] W. Rea, L. Oxley, M. Reale, and J. Brown, "Estimators for long range dependence: an empirical study," *arXiv preprint arXiv:0901.0762*, 2009.

[47] J. Beran, "Statistics for long-memory processes, monographs on statistics and applied probability," *Statitistics Applied Probability Series*, vol. 61, 1994.

[48] M. Li and S. Lim, "Power spectrum of generalized cauchy process," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 219–222, 2010.

[49] ——, "A rigorous derivation of power spectrum of fractional gaussian noise," *Fluctuation and Noise Letters*, vol. 6, no. 04, pp. C33–C36, 2006.

[50] N. J. Kasdin, "Discrete simulation of colored noise and stochastic processes and $1/f^{\alpha}$ power law noise generation," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 802–827, 1995.

[51] M. Li and S. Chen, "Fractional gaussian noise and network traffic modeling," in *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, no. 8. World Scientific and Engineering Academy and Society, 2009.

[52] J. Theiler, "Estimating fractal dimension," *Journal of the Optical Society of America A*, vol. 7, no. 6, pp. 1055–1073, 1990.

[53] T. Kalisky, Y. Ashkenazy, and S. Havlin, "Volatility of fractal and multifractal time series," *Israel Journal of Earth Sciences*, vol. 56, no. 1, pp. 47–56, 2007.

[54] R. Lopes and N. Betrouni, "Fractal and multifractal analysis: a review," *Medical Image Analysis*, vol. 13, no. 4, pp. 634–649, 2009.

[55] Y. Chen, M. Ding, and J. S. Kelso, "Long memory processes ($1/f^{\alpha}$ type) in human coordination," *Physical Review Letters*, vol. 79, no. 22, p. 4501, 1997.

[56] B. Pilgram and D. T. Kaplan, "A comparison of estimators for 1f noise," *Physica D: Nonlinear Phenomena*, vol. 114, no. 1, pp. 108–122, 1998.

[57] A. Diniz, M. L. Wijnants, K. Torre, J. Barreiros, N. Crato, A. M. Bosman, F. Hasselman, R. F. Cox, G. C. Van Orden, and D. Delignières, "Contemporary theories of $1/f$ noise in motor control," *Human Movement Science*, vol. 30, no. 5, pp. 889–905, 2011.

[58] M. J. Cannon, D. B. Percival, D. C. Caccia, G. M. Raymond, and J. B. Bassingthwaighte, "Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series," *Physica A: Statistical Mechanics and its Applications*, vol. 241, no. 3, pp. 606–626, 1997.

[59] R. B. Davies and D. Harte, "Tests for Hurst effect," *Biometrika*, vol. 74, no. 1, pp. 95–101, 1987.

[60] F. Crevecoeur, B. Bollens, C. Detrembleur, and T. Lejeune, "Towards a gold-standard approach to address the presence of long-range auto-correlation in physiological time series," *Journal of Neuroscience Methods*, vol. 192, no. 1, pp. 163–172, 2010.

[61] A. Schaefer, J. S. Brach, S. Perera, and E. Sejdić, "A comparative analysis of spectral exponent estimation techniques for $1/f^{\beta}$ processes with applications to the analysis of stride interval time series," *Journal of Neuroscience Methods*, vol. 222, pp. 118–130, 2014.

[62] B. Audit, E. Bacry, J.-F. Muzy, and A. Arneodo, "Wavelet-based estimators of scaling behavior," *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 2938–2954, 2002.

[63] I. Simonsen, A. Hansen, and O. M. Nes, "Determination of the Hurst exponent by use of wavelet transforms," *Physical Review E*, vol. 58, no. 3, p. 2779, 1998.

[64] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[65] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[66] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. MIT press, 2012.

[67] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," 2007.

[68] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 375–381, 2003.

[69] G. E. Batista and M. C. Monard, "An analysis of four missing data treatment methods for supervised learning," *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 519–533, 2003.

[70] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," in *Artificial Neural Networks ICANN'97*. Springer, 1997, pp. 999–1004.

[71] F. E. Tay and L. Cao, "Application of support vector machines in financial time series forecasting," *Omega*, vol. 29, no. 4, pp. 309–317, 2001.

[72] A. S. Weigend, *Time Series Prediction: Forescasting the Future and Understanding the Past*. Addison-Wesley, 1994, no. 04; QA280, T5.

[73] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT press Cambridge, MA, 1949, vol. 2.

[74] C. M. Hurvich and C.-L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, 1989.

[75] G. Aneiros-Pérez and P. Vieu, "Nonparametric time series prediction: A semi-functional partial linear modeling," *Journal of Multivariate Analysis*, vol. 99, no. 5, pp. 834–857, 2008.

[76] W. G. Cobourn, L. Dolcine, M. French, and M. C. Hubbard, "A comparison of nonlinear regression and neural network models for ground-level ozone forecasting," *Journal of the Air and Waste Management Association*, vol. 50, no. 11, pp. 1999–2009, 2000.

[77] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification.* Wiley-Interscience, 2000.

[78] ——, *Pattern Classification.* John Wiley and Sons, 2012.

[79] M. A. Razi and K. Athappilly, "A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models," *Expert Systems with Applications*, vol. 29, no. 1, pp. 65–74, 2005.

[80] H. J. Motulsky and L. A. Ransnas, "Fitting curves to data using nonlinear regression: a practical and nonmathematical review." *The Federation of American Societies for Experimental Biology journal*, vol. 1, no. 5, pp. 365–374, 1987.

[81] R. J. Oosterbaan and H. Ritzema, "Frequency and regression analysis." *Drainage Principles and Applications.*, no. Ed. 2, pp. 175–223, 1994.

[82] F. D. Foresee and M. T. Hagan, "Gauss-Newton approximation to bayesian learning," in *Neural Networks, 1997., International Conference on*, vol. 3. IEEE, 1997, pp. 1930–1935.

[83] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[84] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[85] G. Bebis and M. Georgiopoulos, "Feed-forward neural networks," *IEEE Potentials*, vol. 13, no. 4, pp. 27–31, 1994.

[86] M. W. Goudreau, C. L. Giles, S. T. Chakradhar, and D. Chen, "First-order versus second-order single-layer recurrent neural networks," *IEEE Transactions on Neural Networks*, vol. 5, no. 3, pp. 511–513, 1994.

[87] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[88] Y. Gao and M. J. Er, "Narmax time series model prediction: feedforward and recurrent fuzzy neural network approaches," *Fuzzy Sets and Systems*, vol. 150, no. 2, pp. 331–350, 2005.

[89] L. Medsker and L. Jain, "Recurrent neural networks," *Design and Applications*, 2001.

[90] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 240–254, 1994.

[91] G. Dorffner, "Neural networks for time series processing," in *Neural Network World*. Citeseer, 1996.

[92] C. L. Giles, S. Lawrence, and A. C. Tsoi, "Noisy time series prediction using recurrent neural networks and grammatical inference," *Machine Learning*, vol. 44, no. 1-2, pp. 161–183, 2001.

[93] S. Cohen, L. Rokach, and O. Maimon, "Decision-tree instance-space decomposition with grouped gain-ratio," *Information Sciences*, vol. 177, no. 17, pp. 3592–3612, 2007.

[94] C. Meek, D. M. Chickering, and D. Heckerman, "Autoregressive tree models for time-series analysis." in *SDM*. Society for Industrial and Applied Mathematics, 2002, pp. 229–244.

[95] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[96] P. J. Brockwell, R. Dahlhaus, and A. A. Trindade, "Modified burg algorithms for multivariate subset autoregression," *Statistica Sinica*, pp. 197–213, 2005.

[97] R. Bos, S. De Waele, and P. M. Broersen, "Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 6, pp. 1289–1294, 2002.

[98] E. Sejdić and L. A. Lipsitz, "Necessity of noise in physiology and medicine," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 2, pp. 459–470, 2013.

[99] E. Sejdić, Y. Fu, A. Pak, J. A. Fairley, and T. Chau, "The effects of rhythmic sensory cues on the temporal dynamics of human gait," *PloS One*, vol. 7, no. 8, p. e43104, 2012.

[100] J. M. Hausdorff, "Gait dynamics, fractals and falls: finding meaning in the stride-to-stride fluctuations of human walking," *Human Movement Science*, vol. 26, no. 4, pp. 555–589, 2007.

[101] J. Kaipust, J. Huisinga, M. Filipi, and N. Stergiou, "Gait variability measures reveal differences between multiple sclerosis patients and healthy controls." *Motor Control*, vol. 16, no. 2, p. 229, 2012.