Tampere University

Matti Molkkari

ADVANCED METHODS IN DETRENDED FLUCTUATION ANALYSIS WITH APPLICATIONS IN COMPUTATIONAL CARDIOLOGY

Engineering and Natural Sciences
Master's Thesis
March 2019

ABSTRACT

Matti Molkkari: Advanced Methods in Detrended Fluctuation Analysis with Applications in Computational Cardiology
Master's Thesis
Tampere University
Science and Engineering, MSc
March 2019

Fractals are ubiquitous in nature. A defining characteristic of fractality is self-similarity; the phenomenon looks similar when observed at multiple scales, which implies the existence of a power law scaling relation. Detrended fluctuation analysis (DFA) is a popular tool for studying these fractal scaling relations. Power laws become linear relationships in logarithmic scales, and conventionally these scaling exponents are determined by simple linear regression in approximately linear regions in doubly logarithmic plots. However, in practice the scaling is hardly ever exact, and its behavior may often vary at different scales.

This thesis extends the fluctuation analysis by introducing robust tools for determining these scaling exponents. A method based on the Kalman smoother is utilized for extracting a whole spectrum of exponents as a function of the scale. The method is parameter-free and resistant to statistical noise, which distinguishes it from prior efforts for determining such local scale exponents. Additionally, an optimization scheme is presented to obtain data-adaptive segmentation of approximately linear regimes. Based on integer linear programming model, the procedure may readily be customized for various purposes. This versatility is demonstrated by applying the method to a group of data to find a common segmentation that is particularly well-suited for machine learning applications.

First, the methods are are employed in exploring the details of the scaling by analyzing simulated data with known scaling properties. These findings provide insight into the interpretation of earlier results.

Second, the methods are applied to the study of heart rate variability. The beating of the heart follows fractal-like patterns, and deviations in these complex variations may be indicative of cardiac diseases. In this context DFA is traditionally performed by extracting two scaling exponents, for short- and long-range correlations, respectively. This has been criticized as an oversimplification, which is corroborated by the results of this thesis. The heart rate exhibits richer fractal-like variability, which becomes apparent in the full scaling spectra. The additional information provided by these methods facilitate improved classification of cardiac conditions.

Keywords: Detrended fluctuation analysis, Kalman filter, Kalman smoother, integer linear programming, heart rate variability, classification

TIIVISTELMÄ

Matti Molkkari: Trendit poistavan fluktuaatioanalyysin edistyneet menetelmät ja niiden sovellukset laskennallisessa kardiologiassa
Diplomityö
Tampereen yliopisto
Teknis-luonnontieteellinen, DI
Maaliskuu 2019

Fraktaaleja esiintyy kaikkialla luonnossa. Fraktaalisuuden ominaispiirre on itsesimilaarisuus, eli ilmiö näyttää samankaltaiselta, kun sitä tarkastellaan useassa eri skaalassa. Tämä johtaa siihen, että ilmiön skaalautuvuus noudattaa potenssilakia. Tällaisia fraktaalisia skaalausrelaatioita voidaan tutkia trendit poistavan fluktuaatioanalyysin avulla (Detrended Fluctuation Analysis, DFA). Logaritminen skaala muuntaa potenssilait lineaarisiksi riippuvuuksiksi, ja tavallisesti skaalauseksponentit määritetään logaritmisista kuvaajista lineaarisen regression avulla. Kuitenkaan skaalautuvuus ei lähes koskaan ole täydellistä, ja se voi myös muuttua eri skaaloilla.

Tämä työ laajentaa fluktuaatioanalyysia esittelemällä paranneltuja menetelmiä näiden skaalauseksponenttien määrittämiseen. Kokonainen spektri skaalauseksponentteja skaalan funktiona määritetään hyödyntämällä Kalman-suodinta. Tämän menetelmän etuja verrattuna aikaisempiin tapoihin määrittää paikallisia skaalauseksponentteja ovat sen parametrivapaa esitys ja vakaus myös kohinaisissa tapauksissa. Lisäksi esitetään lineaariseen kokonaislukuoptimointiin perustuva menetelmä, jonka avulla skaalautuvuudessa voidaan erottaa alueita, jotka noudattavat likimäärin potenssilakia. Tämän mallin muokkaaminen eri tarpeisiin on myös suoraviivaista. Mallia sovelletaan yhteisen segmentaation etsimiseksi datajoukolle, mikä on tarpeen erityisesti koneoppimisen menetelmiä varten.

Esitettyjen menetelmien avulla tutkitaan ensin simuloituja prosesseja, joiden teoreettinen skaalautuminen tunnetaan. Menetelmien mahdollistama yksityiskohtainen analyysi selittää aikaisempia havaintoja DFA:n käyttäytymisessä.

Menetelmiä sovelletaan myös sykevälivaihtelun fraktaalianalyysiin. Terveen sydämen sykeväleissä on fraktaalisia piirteitä, joita eri sairaudet muokkaavat ja hävittävät. Sykevälivaihtelun fraktaalisuutta on perinteisesti kuvattu lyhyen- ja pitkän kantaman skaalauseksponenteilla. Tätä kahden eksponentin mallia on kritisoitu riittämättömäksi, ja tämän työn tulokset vahvistavat tätä näkökulmaa. Skaalauseksponenttispektri paljastaa, että sykevälivaihtelun fraktaalisuus on kahden eksponentin mallia monimuotoisempaa. Esitetyillä menetelmillä saatava lisäinformaatio mahdollistaa aikaisempaa tarkemman sydänsairauksien luokittelun.

Avainsanat: Trendit poistava fluktuaatioanalyysi, Kalman-suodin, lineaarinen kokonaislukuoptimointi, sykevälivaihtelu, luokittelu

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF SYMBOLS AND ABBREVIATIONS

DFA	Detrended Fluctuation Analysis
ECG	Electrocardiogram
FA	Fluctuation analysis
fBm	Fractional Brownian motion
fGn	Fractional Gaussian noise
HRV	Heart Rate Variability
IBI	Interbeat interval
ILP	Integer linear programming
MCC	Matthews correlation coefficient
MF-DFA	Multifractal Detrended Fluctuation Analysis
PSD	Power Spectral Density
RSS	Residual sum of squares
WSS	Weak (or wide)-sense stationarity

$\propto$	Proportional to
$\sim$	Follows the distribution
$\simeq$	Statistical equivalence
$\langle x \rangle$	Arithmetic mean of x
a	Scaling parameter
$\mathbf{A}$	Matrix utilized for computing the weights $G(j, s)$
$\mathbf{A}_k$	State transition matrix
α	Scaling exponent in DFA
$B(t)$	Brownian motion
$B_H(t)$	Fractional Brownian motion with Hurst parameter H
$\mathbf{B}$	Design matrix of ordinary least squares regression
β	Power spectral density scaling exponent
$C(t, s)$	Autocovariance function
$C(\tau)$	Autocovariance function for stationary process
γ	Autocovariance function scaling exponent
$\mathbf{D}$	Step matrix

ϵ	Standard error of the mean
$\mathbb{E}\left[x\right]$	Expectation value of x
f	Frequency
$f_{s,w}(t)$	Trend in window w at scale s
$F(s)$	Fluctuation function
$\tilde{F}(\tilde{s})$	Logarithmic fluctuation function
$\Delta F(s)$	Statistical error estimate of the fluctuation function
$\Delta \tilde{F}(\tilde{s})$	Statistical error estimate of the logarithmic fluctuation function
$F^2_{s,w}$	Squared fluctuations in window w at scale s
$G(j,s)$	Weight function for computing the expectation value of the fluctuation function
$\mathbf{G}_k$	Gain matrix for smoothing distribution
$\Gamma(x)$	The gamma function
H	Hurst parameter
$\mathbf{H}_k$	Measurement model matrix
$\mathbf{I}$	The identity matrix
k	Index in probabilistic state space model
$\mathbf{K}_k$	Optimal Kalman gain
s	DFA window size, another time variable
$\tilde{s}$	Logarithmic DFA window size
$\boldsymbol{m}_0$	Prior for filtering distribution mean
$\boldsymbol{m}_k$	Filtering distribution posterior mean
$\hat{\boldsymbol{m}}_k$	Filtering distribution predicted mean
$\boldsymbol{m}_0^{\mathbf{s}}$	Prior for smoothing distribution mean
$\boldsymbol{m}_k^{\mathbf{s}}$	Smoothing distribution posterior mean
$\hat{\boldsymbol{m}}_k^{\mathbf{s}}$	Smoothing distribution predicted mean
μ	Mean (parameter)
N_s	Number of DFA windows at the scale s
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathbf{P}_0$	Prior for filtering distribution covariance
$\mathbf{P}_k$	Filtering distribution posterior covariance
$\hat{\mathbf{P}}_k$	Filtering distribution predicted covariance
$\mathbf{P}_0^{\mathbf{s}}$	Prior for smoothing distribution covariance
$\mathbf{P}_k^{\mathbf{s}}$	Smoothing distribution posterior covariance

$\hat{\mathbf{P}}_k^{\mathsf{s}}$	Smoothing distribution predicted covariance
q	Multifractality order
$\boldsymbol{q}_k$	Process noise
$\mathbf{Q}_k$	Process noise covariance matrix
$\boldsymbol{r}_k$	Measurement noise
$\mathbf{R}_k$	Measurement noise covariance matrix
$\mathbb{R}$	The set of real numbers
$S(f)$	Power spectral density
$S(\tau)$	Variogram for process with stationary increments
$\mathbf{S}_k$	Innovation covariance matrix
σ^2	Variance
t	Time
T	Maximum index in the state space model
τ	Time lag
$\mathrm{Var}[x]$	Variance of x
$\boldsymbol{v}_k$	Innovation mean
w	DFA window
x_t	Time series
$\boldsymbol{x}_k$	State vector
$\boldsymbol{y}_k$	Measurement vector
Y_t	Random walk time series, integrated profile of the time series

1 INTRODUCTION

Fractals are ubiquitous in nature[1] and even the beating of the heart follows fractal-like patterns [2]. In fact, complex variations in the beat rate are characteristic of a healthy heart, and deviations in this heart rate variability (HRV) may be indicative of cardiac diseases [3]. The study of such fractal correlations in time series was pioneered by Hurst with his research of the yearly discharge of river Nile for the purposes of long-term storage in reservoirs [4]. Hurst's rescaled range analysis (R/S analysis), and other traditional methods, rely on the underlaying assumption that the studied time series is stationary. However, the vast majority of practical time series exhibit non-stationarities, whose exact form is often unknown.

Detrended fluctuation analysis (DFA), originally developed for exploring power law correlations in DNA sequences [5], attempts to overcome these difficulties by proposing a systematic procedure for handling unknown trends in the fluctuation analysis of time series. By computing the mean fluctuations $F(s)$ around least squares polynomial trends at multiple scales s, the method characterizes power law scaling relations $F(s) \propto s^{\alpha}$ by their scaling exponents α [5, 6]. Since its inception, DFA has been widely employed for determining fractal scaling properties and long-range correlations in time series from multiple disciplines, such as physics [7], finance [8], medicine [3] and even music [9]. This thesis provides an up-to-date review of the method and its extensions.

The scaling exponents are conventionally determined by linear regression on a logarithmic scale, which transforms the power laws into linear relationships [6]. In practice the analysis requires subjective consideration, as the scaling is almost never exact and may vary across the scales. This thesis proposes methods for the systematic determination of these scaling exponents without any tunable parameters. This is accomplished by considering a whole spectrum of scaling exponents $\alpha(s)$ as a function of the scale. These spectra are obtained by solving a probabilistic state space model by the Kalman smoother [10, 11]. The prospect of such local scaling exponents has been considered before [12–15], but the earlier approaches are susceptible to noise or rely on adjustable parameters. Additionally, a versatile optimization framework is presented that may be utilized for segmenting the fluctuation functions into regions that exhibit approximate fractal scaling. Other procedures also exist for the detection of such scaling regions by rigorous statistical inference [16, 17]. Here the problem is approached from an optimization viewpoint that is readily customized for different purposes, such as obtaining consistent features for machine learning. The methods introduced in this thesis were briefly presented by the

author at Computing in Cardiology 2018 conference [18].

The majority of the research on the properties of the DFA algorithm itself focuses on its asymptotic behavior and simple linear fitting [6, 19–21]. Therefore the details of the algorithm are first studied by the alpha spectra. However, for applications the main focus will be in studying heart rate variability. While it has been long known that the fractal scaling of the heart rate is altered in diseases [3], it has not yet been fully incorporated into clinical practices [22]. As heart-related mortality is increasingly common, it is a desirable aspiration that these methods could eventually enable earlier detection of potentially life-threatening heart conditions.

The thesis is structured as follows. In Ch. 2 the theoretic prerequisites for understanding the material are discussed, along with a review of the DFA algorithm. In Ch. 3 the theory is applied in constructing novel tools for the robust estimation of the DFA scaling exponent [18]. In Ch. 4 these methods are applied in studying the properties of DFA and analyzing the fractal scaling in heart rate variability. The thesis is wrapped up by the conclusions of Ch. 5.

2 THEORY

This chapter proceeds by a short introduction to the most important concepts for understanding fractality in time series. The following sections describe how random walk theory may be employed for characterizing these fractal correlations, which builds up to detrended fluctuation analysis and culminates in its thorough review. The final sections of this chapter describe the theoretical machinery necessary for understanding the methods presented in this thesis for estimating the scaling exponents of DFA.

2.1 Fractality in time series

The concept of fractals appearing in nature is tangible. Their appreciable beauty is apparent at glance, yet there is hidden elegance in the underlying mathematical machinery. Time series, on the other hand, are collections of numbers representing the observed quantities x_t ordered by the index t. These are usually measurements as a function of discrete time intervals; however, this is not a requirement, and such sequences are traditionally called time series, irrespective of whether the independent variable t corresponds to actual time. The time series may be considered as a realization of an underlying stochastic process. [1, 23]

Therefore, fractality in time series may not be immediately obvious. A defining characteristic of fractals is their self-similarity, and it is the statistical self-similarity that is conserved in fractal time series when scaled by the relation

$$x(at) \simeq a^H x(t), \tag{2.1}$$

where H is the Hurst parameter and $\simeq$ denotes equivalence in the statistical sense, i.e., their probability distributions are equal. Time series exhibiting such fractal scaling relations were first discovered by Hurst, when he was studying variations in the yearly discharge of river Nile [4].

An important concept in time series analysis is stationarity. Intuitively, the series is stationary if its properties, such as the mean and the variance, remain constant in time. Formally, strictly stationary time series has the property that the joint distributions of $(x_{t_1}, \ldots, x_{t_n})$ and $(x_{t_1+\tau}, \ldots, x_{t_n+\tau})$ are equal for all t_i, τ and n. A less prohibiting condition is weak-sense stationarity[1] (WSS), whose definition stipulates first describing the autocovariance

[1] Also known as wide-sense stationarity, and by some authors as second order stationarity.

function $C(t,s)$ of a time series. The general form of the autocovariance function is

$$C(t,s) = \mathbb{E}\left[(x_t - \mu_t)\,(x_s - \mu_s)\right], \tag{2.2}$$

where μ_t and μ_s are the mean values of the process at the times t and s, respectively. The process is WSS if the mean μ is constant and the autocovariance only depends on the lag $\tau = t - s$. Hence the autocovariance function for stationary processes is

$$C(\tau) = \mathbb{E}\left[(x_t - \mu)\,(x_{t+\tau} - \mu)\right] \tag{2.3}$$

for all values of t, which implies that also the variance is constant. [23]

Theoretical models for self-similar time series include fractional Gaussian noise (fGn) and fractional Brownian motion (fBm) [24]. These processes generalize ordinary Gaussian noise and Brownian motion for correlated Gaussian processes. Fractional Brownian motion with Hurst parameter $0 < H < 1$ may be constructed as the moving average of ordinary Brownian motion $\mathrm{d}B(t)$ weighted by the kernel function $(t-s)^{H-1/2}$. Formally this is presented as the fractional integral

$$B_H(t) - B_H(0) = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)}\left\{\int_{-\infty}^{0}\left[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right]\mathrm{d}B(s) + \int_{0}^{t}(t-s)^{H-\frac{1}{2}}\,\mathrm{d}B(s)\right\}. \tag{2.4}$$

The first integral is required to ascertain that the complete history of the process is taken into account, and that the origin is not overemphasized. The normalization by the gamma function ensures that the fractional integration converges appropriately to ordinary repeated integration. The fBm has autocovariance function given by

$$C_H(t,s) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \tag{2.5}$$

where σ^2 is a constant that may be interpreted as the variance of ordinary Gaussian noise. [24]

The increments $B_H(t+\tau) - B_H(t)$ of the fBm, called fractional Gaussian noise, are self-similar with the parameter H and their variance follows the relation

$$S_H(\tau) = \mathbb{E}\left[B_H(t+\tau) - B_H(t)\right]^2 = \sigma^2|\tau|^{2H}, \tag{2.6}$$

which depends only on the lag τ. [24] This quantity $S(\tau)$ describing the expectation value of the squared increments is also known as the variogram for processes with stationary increments [21].

These are continuous-time processes, but in practice they must be sampled at discrete intervals. Therefore, without the loss of generality, the discrete increment process may

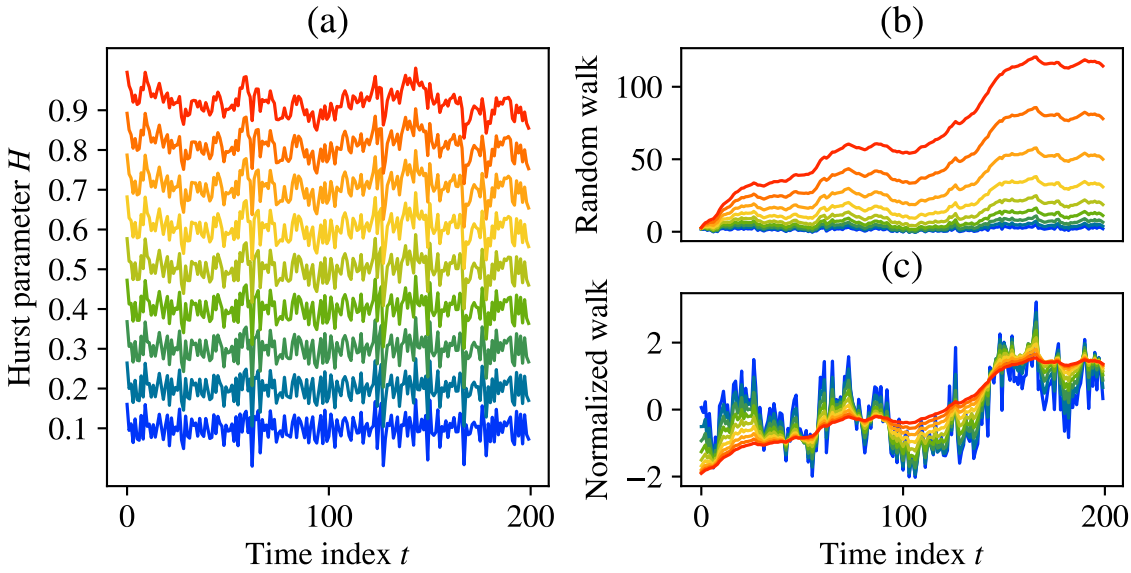

Figure 2.1. Illustration of fractional Gaussian noise and Brownian motion. *(a) Examples of fractional Gaussian noises with various Hurst parameters H and unit variance. The series are generated by the Davies-Harte method utilizing the same random seed. (b) Fractional Brownian motions corresponding to the examples of fGn. (c) The same fBm, except normalized to zero mean and unit variance.*

be defined as $B_H(t) - B_H(t-1)$, leading to the following autocovariance function

$$C_H(\tau) = \frac{\sigma^2}{2}\left(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H}\right). \tag{2.7}$$

Ordinary uncorrelated Gaussian noise, also known as white noise, is recovered for $H = 1/2$. The correlations are negative for $0 < H < 1/2$, implying antipersistent increments. Conversely, for persistent increments with $1/2 < H < 1$ the correlations are positive, and the series $\sum_\tau C_H(\tau)$ diverges, which is conventionally associated with long-range dependence. [21, 24]

Examples of fGn and fBm with different Hurst parameters H are shown in Fig. 2.1. The tendency for similar values is clearly enhanced as the Hurst parameter is increased from the uncorrelated threshold of $H = 1/2$. Likewise the affinity for opposite values is increased as the parameter is decreased. This can be observed as the super- or subdiffusion of the corresponding random walks. Alternatively the diffusion rate may be normalized, in which case persistent walks are characterized by smoother paths, compared to the jaggedness of antipersistent walks.

Another class of fractional noises considered in this thesis are $1/f^\beta$ noises, whose power spectral densities (PSD) obey the asserted scaling as a function of the frequency f. The fGn may be shown to exhibit asymptotic scaling in the frequency domain describable by the exponent β. However, as the fBm is non-stationary process it does not have a unique well-defined description in this realm. [24, 25]

2.2 Random walks and fluctuation analysis

The self-similarity of a signal implies that there exists a quantity of interest $F(s)$ that exhibits power law scaling

$$F(s) \propto s^{\alpha} \tag{2.8}$$

as a function of the scale s, described by the scaling exponent α. Fluctuation analysis (FA) [26] considers, as the name suggests, fluctuations of the signal from its mean and their scaling behavior. Intuitive justification for the method may be derived from random walk theory.

Let us consider a time series x_t consisting of normally distributed independent and uncorrelated random numbers with the mean μ and variance σ^2:

$$x_t \sim \mathcal{N}(\mu, \sigma^2). \tag{2.9}$$

A random walk may be constructed from the series by interpreting the elements x_t as the increments between consecutive steps of the walk. The mean μ must first be subtracted to remove the bias caused by it. The random walk time series Y_t is then obtained as the cumulative sum of the increments:

$$Y_t = \sum_{i=1}^{t} (x_i - \mu)\,. \tag{2.10}$$

The expectation value of this walk is clearly zero

$$\mathbb{E}\left[Y_t\right] = \mathbb{E}\left[\sum_{i=1}^{t} (x_i - \mu)\right] = \sum_{i=1}^{t} \left(\mathbb{E}\left[x_i\right] - \mu\right) = \sum_{i=1}^{t} (\mu - \mu) = 0, \tag{2.11}$$

as expected due to the symmetry about the mean. However, the expectation value for the square of the series is non-zero

$$\mathbb{E}\left[Y_t^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{t} (x_i - \mu)\right)^2\right] = t\sigma^2, \tag{2.12}$$

which is equal to the variance as the mean has been subtracted[2]. This yields the well known result for the root mean square displacement of an uncorrelated random walker

$$\sqrt{\mathbb{E}\left[Y_t^2\right]} \propto t^{\frac{1}{2}}, \tag{2.13}$$

which is a power law scaling relation with $\alpha = 1/2$. However, if the increments $x_i - \mu$ are correlated, then these consecutive steps are more likely to be towards the same direction, resulting in hastened diffusion of the walker indicated by $\alpha > 1/2$. Conversely,

[2] Computing this is more tedious, but utilizing the rule for the sum of two independent Gaussian variables yields the result by induction.

the enhanced probability for opposite steps with anticorrelated increments suppresses the progress of the walker and leads to scaling exponents $\alpha < 1/2$. This effect can be observed in Fig. 2.1 where the Hurst parameter H is equal to this exponent α.

In fluctuation analysis the time series is similarly interpreted as the increments of a random walk [26]. The analysis proceeds by constructing a fluctuation function $F(s)$, which corresponds to the root mean square displacement of Eq. (2.13). In this context the scale s is understood as the number of steps taken by the random walker. The resulting fluctuation function

$$F(s) = \sqrt{\mathbb{E}\left[(Y_{t+s} - Y_t)^2\right]} \tag{2.14}$$

is readily interpreted as the root mean square displacement at the scale s. The expectation value is usually computed by dividing the time series into non-overlapping windows of length s, over which the squared differences are averaged [27].

The analysis so far has assumed that the time series is stationary, so that the expectation value in Eq. (2.14) is independent of t. In practice trends are usually present in the time series, which may result in spurious detection of long-range correlations. Often the trends and their characteristic time scales are unknown, further complicating the analysis. Detrended fluctuation analysis [5] attempts to overcome this problem by detrending the integrated profile Y_t in each window. This detrending is usually performed by computing a least squares fit of a low order polynomial to the profile, which is then subtracted.

More formally, the detrending procedure within each window w of length s yields the mean squared fluctuations $F^2_{s,w}$ defined by

$$F^2_{s,w} = \frac{1}{s} \sum_{t \in w} [Y_t - f_{s,w}(t)]^2, \tag{2.15}$$

where $f_{s,w}$ is the trend within that particular window. These squared fluctuations correspond to the variance if the detrended profiles have means equal to zero within the windows. The zero mean condition is automatically fulfilled by least squares detrending polynomials. The variance directly characterizes the squared displacement of random walkers. Therefore the connection to random walks is preserved by the fluctuation function defined as

$$F(s) = \sqrt{\mathbb{E}\left[F^2_{s,w}\right]}, \tag{2.16}$$

where the expectation value is obtained by averaging over all the windows for each scale separately. The purpose of the detrending is to ensure that this expectation value is independent of the window w, which will be further discussed in Sec. 2.3.2.

At this point a brief digression is in order to discuss the computation of the average scaling behavior of many samples. A naïve approach would be to simply calculate the fluctuation function for each sample, as per Eq. (2.16), and then compute the average

from these. However, the result will be biased with the importance of larger fluctuations diminished. This is because the expectation value is supposed to be taken from the squared fluctuations. Taking the square root is the final operation, so that the fluctuations are expressed in the same units as the original time series. Therefore, the mathematically correct method is to compute the expectation value in Eq. (2.16) with the squared fluctuations of Eq. (2.15) from *all* the samples. The matter is complicated by the implicit assumption that each sample is drawn from the same process with the same parameters. In reality there almost always exists individual variability between samples, and samples consisting of values of larger magnitude would dominate the calculation, as would longer samples. Consequently some sort of normalization technique may be required when dealing with practical data.

The final step in the analysis consists of determining the scaling exponent α from the fluctuation function that has been computed at multiple scales. Power laws of the form described by Eq. (2.8) may be determined from doubly logarithmic plots of the fluctuation function versus the scale, as any power law scaling would become linear when observed on logarithmic scale.

These steps of the algorithm—computation of the fluctuation function, detrending, and in particular robust determination of the scaling exponent—will be discussed in more detail in the forthcoming sections of the thesis.

2.3 Detrended fluctuation analysis

Detrended fluctuation analysis was originally developed as a response to criticism citing non-stationarities as the possible cause for the long-range correlations discovered in DNA sequences [5]. The traditional DFA formulation [5, 6] is described below, complemented by simple statistical error measure for the fluctuation function [18].

The algorithm proceeds by constructing the integrated profile Y_t, analogous to the random walk time series of Eq. (2.10), as the cumulative sum

$$Y_t = \sum_{i=1}^{t} (x_t - \langle x \rangle) \qquad\qquad t = 1, \ldots, N, \tag{2.17}$$

where $\langle x \rangle$ denotes the mean of the series. The profile Y_t is then divided into N_s contiguous non-overlapping windows at various scales s. For every window size s the local trend $f_{s,w}(t)$ is determined in each window w by least squares polynomial regression. The algorithm is conventionally denoted by DFAn, where n is the degree of the detrending polynomial [6].

The squared fluctuations $F^2_{s,w}$, corresponding to the variance, within each window are given by Eq. (2.15). For notational convenience, let μ_s and ϵ_s denote the mean and its

standard error, respectively, of the squared fluctuations over all windows of size s

$$\mu_s = \left\langle F^2_{s,w} \right\rangle \qquad\qquad \epsilon_s = \sqrt{\frac{\mathrm{Var}[F^2_{s,w}]}{N_s}}. \tag{2.18}$$

The fluctuation function $F(s)$ and its error estimate[3] $\Delta F(s)$ may then be expressed as

$$F(s) = \sqrt{\mu_s} \qquad\qquad \Delta F(s) \approx \frac{\epsilon_s}{2\sqrt{\mu_s}}. \tag{2.19}$$

It should be noted that this error estimate only quantifies the statistical error in determining the mean from finite sample size, and is disconnected from any possible systematic errors present in the algorithm. Power laws $F(s) \propto s^\alpha$ are conveniently determined from the corresponding logarithmic quantities, and these shall be denoted by tildes:

$$\tilde{F}(\tilde{s}) = \log_{10} \sqrt{\mu_s} \qquad \Delta\tilde{F}(\tilde{s}) \approx \frac{\epsilon_s}{\mu_s \ln 100} \qquad \tilde{s} = \log_{10} s. \tag{2.20}$$

The DFA algorithm is graphically illustrated in Fig. 2.2.

2.3.1 Fluctuation function

Conventionally the fluctuation function is calculated by utilizing non-overlapping windows when computing the squared fluctuations. This is problematic particularly for larger window sizes, as the number of available windows is restricted by the length of the time series. Allowing the windows to overlap permits statistics with larger sample size, but introduces new types of errors: the squared fluctuations are not independent from each other anymore, and the outermost elements of the time series do not contribute equally to the result.

Another complication arises from the reality that most window sizes do not exactly divide the length of the time series. Therefore it has become customary to repeat the windowing procedure from the other end of the record as well, in order not to disregard any data. However, in doing so correlations are introduced amongst the squared fluctuations. The largest window size is not recommended to exceed $N/4$ when utilizing this bidirectional windowing procedure. [6] It is advisable to inspect the fluctuation function for outliers and judge its statistical accuracy on case-by-case basis.

It is, however, possible to derive analytical expression for the expectation value of the square of the fluctuation function that does not depend on any windowing procedure. For stationary signals this expectation value may be expressed as a weighted sum of the autocovariance function[20, 21], and for non-stationary processes with stationary increments as a weighted sum of the variogram[21]. Following the approach in Ref. [21], the

[3] The propagation of error Δx through function $f(x)$ is approximated by

$$\Delta f(x) \approx \left| \frac{\partial f(x)}{\partial x} \right| \Delta x$$

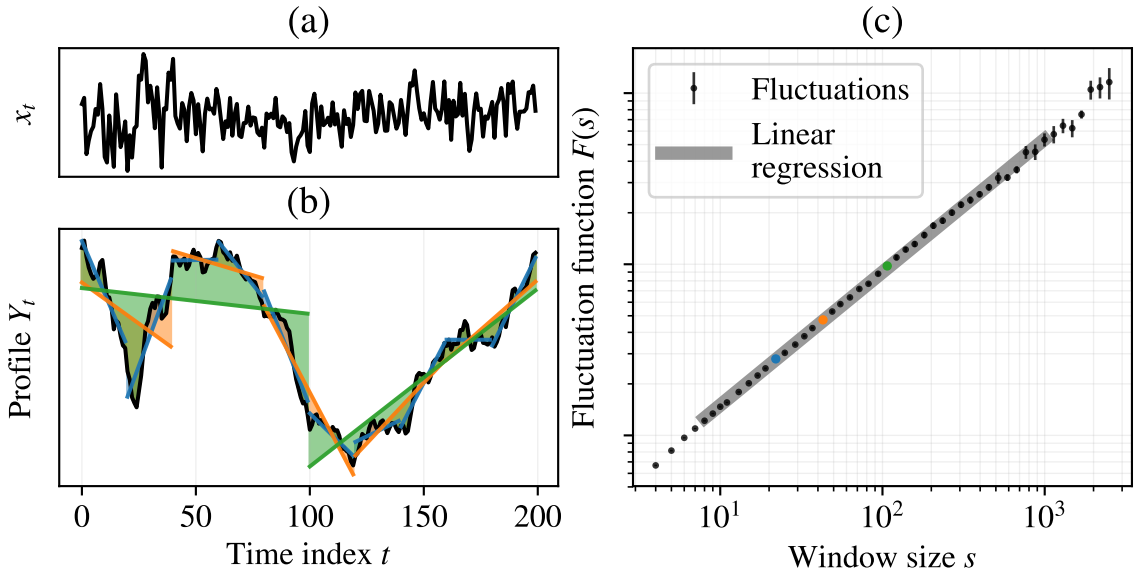

Figure 2.2. Illustration of the traditional DFA method. *(a) An example of a time series consisting of fractional Gaussian noise with Hurst parameter $H = 0.9$. (b) The integrated profile of the example time series. Trends are determined in non-overlapping windows at various scales by least squares polynomial fitting. In this figure linear detrending of DFA1 is illustrated with window sizes of 20 (blue), 40 (orange) and 100 (green). Fluctuations in each window are defined as the mean squared difference from the local trend, which corresponds to the variance. (c) The fluctuation function at each scale s is the square root of the expectation value of the squared fluctuations at that scale. The scaling exponent α may be estimated as the slope of the fluctuation function on a doubly logarithmic plot by simple linear regression. As explained in the text, the fluctuation function exhibits bias at the shorter scales and becomes statistically unreliable at larger scales, which necessitates judicious decision about suitable fitting ranges.*

weight function $G(j, s)$ is compactly written as the sum of matrix elements

$$G(j,s) = \sum_{k=1}^{s-j} a_{k,k+j}, \tag{2.21}$$

where $a_{k,k+j}$ are the elements of the following matrix

$$\mathbf{A} = \mathbf{D}^\top \Big(\mathbf{I} - \mathbf{B}^\top \big(\mathbf{B}\mathbf{B}^\top \big)^{-1} \mathbf{B} \Big) \mathbf{D} \tag{2.22}$$

with $\mathbf{I} \in \mathbb{R}^{s \times s}$ as the identity matrix. The elements of matrix $\mathbf{D} \in \mathbb{R}^{s \times s}$ are defined as

$$d_{i,j} = \begin{cases} 1 & i \leq j \\ 0 & \text{otherwise} \end{cases}, \tag{2.23}$$

which causes left multiplying a column vector by $\mathbf{D}$ to yield a row vector of cumulative sums. Detrending by n-th degree polynomials is accomplished by the design matrix $\mathbf{B} \in$

$\mathbb{R}^{n+1\times s}$ of ordinary least squares regression. That is, the k-th row of $\mathbf{B}$ is defined as

$$B_k = \begin{bmatrix} 1^{k-1} & 2^{k-1} & \dots & s^{k-1} \end{bmatrix}. \tag{2.24}$$

Defining $\boldsymbol{X}_{s,w} \in \mathbb{R}^s$ to be a column vector containing the elements of the time series in window w at the scale s allows the squared fluctuations to be written as

$$F^2_{s,w} = \frac{1}{s}\boldsymbol{X}^\top_{s,w}\mathbf{A}\boldsymbol{X}_{s,w} \tag{2.25}$$

$$= \frac{1}{s}\sum_{k=1}^{s}\sum_{j=1}^{s} a_{k,j}x_{w+k}x_{w+j} \tag{2.26}$$

$$= -\frac{1}{2s}\sum_{k=1}^{s}\sum_{j=1}^{s} a_{k,j}\left(x_{w+k} - x_{w+j}\right)^2. \tag{2.27}$$

Taking the expectation value of Eq. (2.26) yields

$$\mathbb{E}\left[F^2_{s,w}\right] = \frac{1}{s}\sum_{k=1}^{s}\sum_{j=1}^{s} a_{k,j}C(w+k, w+j), \tag{2.28}$$

where $C(w+k, w+j)$ is the autocovariance function. If the process is assumed to be stationary, then

$$C(w+k, w+j) = C(0, w+j-(w+k)) \coloneqq C(\tau) \tag{2.29}$$

with $\tau = j - k$ being the lag. Thus the expectation value is independent of the window w and may be written as

$$\mathbb{E}\left[F^2(s)\right] = \frac{1}{s}\left[C(0)G(0,s) + 2\sum_{j=1}^{s-1} G(j,s)C(j)\right]. \tag{2.30}$$

Alternatively, applying the expectation operator to Eq. (2.27) and assuming the stationarity of the increments gives

$$\mathbb{E}\left[F^2(s)\right] = -\frac{1}{s}\sum_{j=1}^{s-1} G(j,s)S(j), \tag{2.31}$$

where $S(j)$ is the variogram. [21]

These relations may also be utilized for computing the expected theoretical fluctuation functions for processes with known autocovariance function or variogram, e.g. for fractional Brownian motion. Furthermore, these expressions facilitate proving the correct asymptotic scaling of the fluctuation function in these cases [21]. In this thesis these analytical methods are utilized in studying the detailed scaling of fGn and fBm in Ch. 4.1, along with comparison to numerical experiments employing the proposed new methods.

2.3.2 Detrending

Traditional fluctuation analysis fails to find intrinsic scaling relations in the presence of non-stationarities, as they would result in an asymptotic scaling exponent of unity. Therefore, detrending may be considered the most important feature of DFA, as it is precisely the aspect that sets the method apart from earlier approaches, and permits meaningful fractal correlation analysis of non-stationary time series. In applications it is important to distinguish the difference between external and intrinsic non-stationarities. [28] For example, the diffusion-like drift in fractional Brownian motion is intrinsic to the process, whereas the variations in heart rate due to different activities can be considered as responses to external stimuli. Contrary to the criticism about the failure of DFA to remove the drift in fBm [29], it may instead be considered as a crucial advantage of the method to correctly characterize this intrinsically non-stationary process [28].

Conventionally the detrending is performed by ordinary least squares fits of polynomials to the segmented profile of the time series. Zeroth order detrending corresponds to the removal of the mean in Eq. (2.15), which then reduces to ordinary variance. This DFA0 formulation is therefore essentially equivalent to the conventional fluctuation analysis with the same limitations[4].

Linear detrending has been shown to be sufficient for obtaining the correct scaling exponents for fBm time series [21, 28]. Higher orders of detrending retain this correct scaling asymptotically with large window sizes [30]. The minimum detectable DFA scaling exponent is always zero, but the maximum α_{max} depends on the degree n of the detrending polynomial and is given by [30]

$$\alpha_{\text{max}} = n + 1. \tag{2.32}$$

As the detrending is performed for the integrated profile of the time series, DFAn corresponds to removing polynomial trends of degree $n - 1$ in the original time series. It can be shown that this piecewise detrending does indeed recover the scaling exponent that would characterize the time series in the absence of the external polynomial trend [21].

It is enthralling to contemplate how exactly does the detrending process allow non-stationarities to be properly considered. It can be shown that a sufficient, but probably not necessary, condition for the correct scaling of the fluctuation function is that the windowed variances $F^2_{s,w}$ are equal in expectation value at each scale. Therefore, the purpose of the detrending is to make each window statistically equivalent. The detrending procedure in Eq. (2.15) is additive, so it can be hypothesized that the underlying assumption in DFA is that the removal of superimposed least-squares polynomial trends is sufficient to ascertain this statistical equivalence. [28]

[4] The variance is, however, a more robust estimator for the root mean square displacement of a random walker than the literal displacement of Eq. (2.13), which allows the analysis to be performed on shorter time series.

While necessary, the detrending also introduces artifacts to the analysis. The most pronounced effects arise in the smallest windows, where the fluctuations are underestimated. If the time series exhibits scaling then this bias may be corrected. A modified fluctuation function $F_{\mathsf{mod}}(s)$ with the bias corrected may be computed by

$$F^2_{\mathsf{mod}}(s) = \frac{F^2(s)}{K^2_\alpha(s)}, \tag{2.33}$$

where $K^2_\alpha(s)$ is the correction function. Further complications arise as this function depends on the scaling exponent α. The correction term may be computed by assuming that the scaling relation $F^2(s_\infty) \propto s_\infty^{2\alpha}$ holds for sufficiently large window size s_∞. With this assumption the correct scaling is restored by

$$K^2_\alpha(s) = \frac{\mathbb{E}\left[F^2(s)\right] s_\infty^{2\alpha}}{\mathbb{E}\left[F^2(s_\infty)\right] s^{2\alpha}}. \tag{2.34}$$

If the theoretical autocovariance function or variogram is known for the underlying process, then these expectation values may be computed by Eq. (2.30) or (2.31), depending whether the process is stationary or non-stationary with stationary increments. [21] If the analytical formulation for these quantities is not available, then these expectation values may be approximated by ensemble averages of shuffled time series [6]. Attempts have also been mode to modify the detrended variance computation in such a manner as to remain unbiased [31].

The detrending is conventionally performed by polynomials, but other approaches have been studied as well. Recently detrending moving average (DMA) has been established as a good alternative to the polynomial DFA [32]. Empirical mode decomposition (EMD) has also been employed for the detrending process [33]. The detrending step of the algorithm is essential for intrinsic non-stationarities, but the removal of external trends may be performed before the fluctuation analysis as well. This is particularly fruitful if the trend is known *a priori*. Even in the case of unknown trends this may sometimes be useful. An interesting approach is offered in Ref. [34], where the learning capabilities of echo state networks (ESN) are leveraged to distinguish the more predictable components of the time series as trends.

2.3.3 Scaling exponent

Once the fluctuation function has been computed, its scaling properties are quantified by scaling exponents. Self-similar processes exhibit power law scaling $F(s) \propto s^\alpha$ as a function of the scale s. For many processes, including fGn and fBm, the fluctuation function only approaches the exact scaling asymptotically in the large scale limit [21]. In more complex processes, and particularly with real experimental data, the scaling may be limited to certain ranges of scales. Different scales may also display distinct scaling, and the crossover scales over which the behavior changes may yield information about the underlying dynamical process.

Further complications arise from finite size effects present in the fluctuation function, as the sample size is very small for the largest scales, resulting in unreliable statistics with large confidence intervals. The detrending process itself is also found to introduce bias to the fluctuation function, especially at the smaller scales. Conventionally the scaling exponent is determined from a doubly logarithmic plot of the fluctuation function, which transforms the power law into linear relationship and permits the application of simple linear regression, which was illustrated in Fig. 2.2. Furthermore, this least-squares fitting may be misleading for determining power law behavior [35]. As a result of the aforementioned intricacies, the determination of the scaling exponent is a non-trivial task.

Many techniques exist for the detection of crossover points and the subsequent determination of approximately linear regimes in the fluctuation function. The most rigorous methods rely on piecewise linear regression and statistical inference to determine the number and location of the crossover points [16, 17]. Another particularly intriguing strategy involves expressing the fluctuation function as a linear combination of power law functions [36]. This thesis examines an integer linear programming scheme for determining the optimal segmentation, which will be presented in Sec. 3.2.

The problem of determining strictly scaling regions in the fluctuation function may be circumvented by considering a whole spectrum of exponents $\alpha(s)$ as a function of the scale. As logarithmically transformed power law relations are straight lines, the local scaling exponent could be estimated as the derivative of the logarithmic fluctuation function [12, 14]. However, direct numerical differentiation can be very noisy, particularly for shorter time series whose fluctuation functions suffer from statistical uncertainties due to the small sample size. Smoother local estimates could be acquired by, for example, utilizing short linear fitting windows moving over the scales [15], or applying some filtering, such as the alpha-beta filter [13]. The problem with these methods is that they rely on hand-picked parameters, usually chosen in an *ad hoc* manner. This thesis considers a parameter-free approach based on the Kalman smoother, which will be introduced in the succeeding chapter.

Different applications of DFA have also adopted specialized conventions for the determination of the scaling exponents. Particularly in the context of heart rate variability, two scaling exponents, one for short- and another for long-range correlations, are usually extracted from the fluctuation function [2]. However, the extent of these ranges is subjective, and different data sets may suggest conflicting crossover regions, leading to criticism about the validity of this traditional two-range model [13–15].

The DFA scaling exponent is also related to many other quantities describing similar scaling relations in other domains. In Hurst's seminal work on long-range correlated processes, the increment process is characterized by an exponent obtained from his rescaled range analysis [4]. This Hurst parameter H and the associated stochastic processes, fGn and fBm, were rigorously studied by Mandelbrot and van Ness [24]. For fGn, the Hurst parameter is asymptotically equal to the DFA scaling exponent. For fBm, the relation $\alpha = H + 1$ holds asymptotically [21].

Table 2.1. *Interpretation of the DFA scaling exponent.*

Scaling exponent	Interpretation	Stationarity
$0 < \alpha < 1/2$	Anticorrelated	Stationary
$\alpha = 1/2$	White noise	
$1/2 < \alpha < 1$	Correlated	
$\alpha = 1$	Pink noise	
$1 < \alpha < 1 1/2$	Anticorrelated increments	Non-stationary, stationary increments
$\alpha = 1 1/2$	Brownian motion	
$1 1/2 < \alpha < 2$	Correlated increments	

Fourier methods are also common in time series analysis. Specifically, scaling properties may be discovered in the power spectral density (PSD), $S(f) \propto f^{-\beta}$, which describes the amount of power[5] associated with the frequency f. The PSD scaling exponent β is related to the DFA scaling exponent by $\beta = 2\alpha - 1$. [25, 30] Comparison of PSD and DFA results is further complicated by non-trivial distortion in the relationship between the DFA scales s and PSD frequencies f. As the order of the DFA is increased, the frequencies f correspond to progressively larger scales s. Similarly signals with smaller PSD scaling exponents β shift the frequencies f towards increasing scales s as well. [30]

For wide-sense stationary processes the Wiener-Khinchin theorem holds, which tells that the autocovariance function and power spectral density are Fourier transform pairs [37]. This implies that the autocovariance function $C(\tau)$ also exhibits scaling $C(\tau) \propto \tau^{-\gamma}$ with $\gamma = 1 - \beta$. Therefore the DFA scaling exponent is related to these other exponents by

$$\alpha = \frac{1}{2}(\beta + 1) \tag{2.35}$$

$$\alpha = 1 - \frac{1}{2}\gamma. \tag{2.36}$$

As the DFA method is based on random walk theory, it is intrinsically linked to diffusion. The scaling exponent is essentially equal to the anomalous diffusion exponent, except that instead of measuring standard deviation, the diffusion exponent characterizes variance, and hence their usual definitions differ by a factor of two [38].

The cumulative summation of Eq. (2.10) may be repeated for the profile Y_t, in which case the original profile becomes the increment time series. This increases the scaling exponent by one each time the time series is "integrated" in such manner. Conversely, a "differentiated" time series may be constructed by $x'_t = x_{t+1} - x_t$, which reduces the exponent by one. [27] These techniques may be utilized for transforming the time series to have scaling exponents in the detectable range of $[0, n + 1]$.

In light of this discussion, the interpretation of the scaling exponent is summarized in Table

[5] The squared value of the time series, which corresponds to actual power density in certain physical systems

2.1. One must be careful in drawing too strong conclusions from the scaling exponent, as these interpretations are only asymptotically valid for strictly scaling time series, such as fGn, fBm or $1/f^{\beta}$ noises. In practice these are hardly ever exact models for the studied processes, and therefore one cannot conclude that a measured exponent of, e.g. 1.2 implies that the process is fBm with Hurst parameter $H = 0.2$. Another often encountered process is pink noise whose PSD follows $1/f$ relation. By Eq. (2.35) this implies $\alpha = 1$. The cumulative summation procedure can be further repeated for even higher values of α. The qualitative interpretation remains the same, that is, for each integral interval the lower and upper halves correspond to originally anticorrelated or correlated increments, respectively.

2.3.4 Multifractality

The time series may also exhibit more complex behavior than what can be described by merely considering the correlations at different scales. The scaling may also vary throughout the time series, even in a very complex manner across different interwoven subsets. This kind of multifractality can be understood to be a consequence of broad probability density function for the values of the time series, or due to different correlations amongst the smaller and larger fluctuations. Conventional tools of multifractal analysis, such as the partition function formalism, suffer from complications arising from non-stationarities. [39, 40]

Multifractal detrended fluctuation analysis (MF-DFA) was developed to tackle these problems [40]. Ordinary monofractal DFA estimates the fluctuation function directly from the mean of the variances of Eq. (2.15). For MF-DFA the averaging procedure is altered to obtain separate fluctuation functions for smaller and larger fluctuations. This is accomplished by raising these variances to different powers before calculating the mean, and then transforming the mean back to the correct dimensions by the corresponding inverse power[6]. Thus the fluctuation function is computed by

$$F_q(s) = \left(\mathbb{E} \left[\left(F_{s,w}^2 \right)^{\frac{q}{2}} \right] \right)^{\frac{1}{q}}, \tag{2.37}$$

where q is the multifractality order. [40] Looking at the expression, it is clear that for positive q, the larger the multifractality order, the bigger the contribution of larger variances $F_{s,w}^2$ is to the mean. Conversely for negative values of q, smaller variances become more and more dominant with decreasing multifractality order[7]. Setting $q = 2$ recovers the ordinary monofractal DFA. For $q = 0$ the outer exponent diverges, and the following

[6] One must remember that the $F_{s,w}^2$ are variances, hence the imbalance by a factor of 2 in the denominator.

[7] As a practical observation there are usually numerical issues for negative values of q in the smallest windows, where the variances are the smallest and the number of windows is the highest. Particularly chains of (almost) equal values in the time series are problematic.

logarithmic averaging procedure must be employed [40]

$$F_0(s) = \exp\left[\frac{1}{2}\mathbb{E}\left[\ln\left(F_{s,w}^2\right)\right]\right]. \tag{2.38}$$

To facilitate the computation of statistical error estimates for this multifractal formalism, the following auxiliary variables are introduced

$$\xi_{q,s,w} = \begin{cases} \left(F_{s,w}^2\right)^{\frac{q}{2}} & q \neq 0 \\ \ln\left(F_{s,w}^2\right) & q = 0 \end{cases}. \tag{2.39}$$

Let $\mu_{q,s}$ and $\epsilon_{q,s}$ denote the mean and its standard error, respectively, of these variables over all windows w of length s

$$\mu_{q,s} = \langle \xi_{q,s,w} \rangle \qquad\qquad \epsilon_{q,s} = \sqrt{\frac{\mathrm{Var}[\xi_{q,s,w}]}{N_s}}. \tag{2.40}$$

Analogously to ordinary DFA, the multifractal fluctuation functions and their error estimates may then be expressed as

$$F_q(s) = \begin{cases} \left(\mu_{q,s}\right)^{\frac{1}{q}} & q \neq 0 \\ e^{\frac{1}{2}\mu_{q,s}} & q = 0 \end{cases} \qquad \Delta F_q(s) \approx \begin{cases} \frac{1}{q}\left(\mu_{q,s}\right)^{\frac{1}{q}-1}\epsilon_{q,s} & q \neq 0 \\ \frac{1}{2}e^{\frac{1}{2}\mu_{q,s}}\epsilon_{q,s} & q = 0 \end{cases}. \tag{2.41}$$

By considering each value of q separately, the scaling exponent may be determined by the same techniques as in monofractal DFA. However, the additional dependence on q allows more comprehensive analysis by considering for example the surfaces spanned by the scaling exponent $\alpha(s, q)$ as a function of the scale and the multifractal order [41]. The MF-DFA algorithm utilized in this thesis is summarized in Algorithm 2.1.

2.4 Bayesian inference

Let us consider a probabilistic state space model of a system, described by its hidden state $\boldsymbol{x}_k \in \mathbb{R}^n$ at the time step k, following the probability distribution

$$\boldsymbol{x}_k \sim p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}). \tag{2.47}$$

The system is Markovian in the sense that its evolution only depends on the previous state. A measurement of the system yields the results $\boldsymbol{y}_k \in \mathbb{R}^m$ described by the probability distribution

$$\boldsymbol{y}_k \sim p(\boldsymbol{y}_k|\boldsymbol{x}_k). \tag{2.48}$$

However, the quantity of interest is the probability distribution of the hidden state of the system, given the measurements, which can be calculated by Bayes' rule

$$p(\boldsymbol{x}_k|\boldsymbol{y}_k) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k)}{p(\boldsymbol{y}_k)}. \tag{2.49}$$

Algorithm 2.1 Multifractal detrended fluctuation analysis.

1. Let x_t be a time series with N elements.
2. Calculate the profile Y_t of the time series

$$Y_t = \sum_{i=1}^{t} (x_i - \langle x \rangle) \qquad t = 1, \ldots, N, \tag{2.42}$$

   where $\langle x \rangle$ is the mean of the time series.
3. Windowing
   (a) Decide a suitable set of different window sizes $\mathcal{S}$.
   (b) For each scale $s \in \mathcal{S}$ divide the profile Y_t into $\lfloor N/s \rfloor$ contiguous non-overlapping segments $\mathcal{W}_s$ of length s, starting from the first element of the profile. If the window size s does not exactly divide the length N of the time series, repeat the segmentation from the opposite end of the profile as well.
4. Detrending
   (a) Decide a suitable type of detrending function.
   (b) In each window $w \in \mathcal{W}_s$ for all scales $s \in \mathcal{S}$ perform a least squares fit of the detrending function $f_{s,w}(t)$ to the profile Y_t and calculate the detrended variance $F^2_{s,w}$ in the segment:

$$F^2_{s,w} = \frac{1}{s} \sum_{t \in w} \left[Y_t - f_{s,w}(t) \right]^2 \tag{2.43}$$

5. Calculate the mean fluctuations with the desired multifractal orders q.
   (a) Let

$$\xi_{q,s,w} = \begin{cases} \left(F^2_{s,w}\right)^{\frac{q}{2}} & q \neq 0 \\ \ln\left(F^2_{s,w}\right) & q = 0 \end{cases} \tag{2.44}$$

   (b) Calculate the means $\mu_{q,s}$ and their standard errors $\epsilon_{q,s}$

$$\mu_{q,s} = \langle \xi_{q,s,w} \rangle_{w \in \mathcal{W}_s} \qquad \epsilon_{q,s} = \sqrt{\frac{\mathrm{Var}[\xi_{q,s,w}]_{w \in \mathcal{W}_s}}{|\mathcal{W}_s|}}, \tag{2.45}$$

   where the mean and the variance are calculated over all segments $w \in \mathcal{W}_s$ for each scale s, and $|\mathcal{W}_s|$ is the number of elements in $\mathcal{W}_s$.
   (c) Compute the mean fluctuations $F_q(s)$ and the error estimates $\Delta F_q(s)$:

$$F_q(s) = \begin{cases} \left(\mu_{q,s}\right)^{\frac{1}{q}} & q \neq 0 \\ e^{\frac{1}{2}\mu_{q,s}} & q = 0 \end{cases} \qquad \Delta F_q(s) \approx \begin{cases} \frac{1}{q}\left(\mu_{q,s}\right)^{\frac{1}{q}-1} \epsilon_{q,s} & q \neq 0 \\ \frac{1}{2} e^{\frac{1}{2}\mu_{q,s}} \epsilon_{q,s} & q = 0 \end{cases} \tag{2.46}$$

Bayesian filtering attempts to estimate the state of the system at time step k, given all the past measurements up to the current time step $k \leq T$, to produce the posterior distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$. Bayesian smoothing attempts to estimate the state of the system at time step k, given *all* the measurements[8], to produce the posterior distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:T})$. The direct application of the Bayes' rule for calculating these distributions, especially for real-time applications, may be nigh intractable for general models. [42] The Kalman filter provides a computationally light recursive solution to linear models disturbed by normally distributed noise [10].

2.4.1 Kalman filter

The notation and exposition in this section closely follows that of Ref. [42]. The Kalman filter[10] assumes linear models for the state evolution and measurements, disturbed by Gaussian noise. Thus the state of the system $\boldsymbol{x}_k$ evolves from the previous state by

$$\boldsymbol{x}_k = \mathbf{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{q}_{k-1}, \tag{2.50}$$

where $\mathbf{A}_{k-1} \in \mathbb{R}^{n\times n}$ is the state transition matrix, and $\boldsymbol{q}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ describes the process noise with covariance matrix $\mathbf{Q}_{k-1} \in \mathbb{R}^{n\times n}$. The measurement model is similarly linear, with the hidden state $\boldsymbol{x}_k$ yielding the measurement $\boldsymbol{y}_k$ given by

$$\boldsymbol{y}_k = \mathbf{H}_k\boldsymbol{x}_k + \boldsymbol{r}_k, \tag{2.51}$$

where $\mathbf{H}_k \in \mathbb{R}^{m\times n}$ is the measurement model matrix, and $\boldsymbol{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ describes the measurement noise with covariance matrix $\mathbf{R}_k \in \mathbb{R}^{m\times m}$. The linearity of the equations ensures that given a normally distributed prior distribution $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{m}_0, \mathbf{P}_0)$, the filtering and smoothing posterior distributions remain Gaussian.

The Kalman filter is a recursive closed form solution for the filtering distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \mathcal{N}(\boldsymbol{m}_k, \mathbf{P}_k)$, which is outlined below [10, 42]. Let us assume a normally distributed previous state $\boldsymbol{x}_{k-1} \sim \mathcal{N}(\boldsymbol{m}_{k-1}, \mathbf{P}_{k-1})$ with mean $\boldsymbol{m}_{k-1}$ and covariance $\mathbf{P}_{k-1}$. The predicted current state of the system $\hat{\boldsymbol{x}}_k \sim p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{1:k-1}) = \mathcal{N}(\hat{\boldsymbol{m}}_k, \hat{\mathbf{P}}_k)$ can then be obtained by propagating the previous distribution through the state evolution model described by Eq. (2.50). The predicted mean $\hat{\boldsymbol{m}}_k$ and covariance $\hat{\mathbf{P}}_k$ of the current state are then given by

$$\hat{\boldsymbol{m}}_k = \mathbf{A}_{k-1}\boldsymbol{m}_{k-1} \tag{2.52}$$

$$\hat{\mathbf{P}}_k = \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^\top + \mathbf{Q}_{k-1}. \tag{2.53}$$

However, this prediction does not yet take into account the measurement $\boldsymbol{y}_k$ at the current step.

The Kalman filter attempts to minimize the mean squared error of the estimate, which

[8] Obvious limitations may apply to procuring measurements from the future in real-time applications.

is equivalent to minimizing the trace of the posterior distribution covariance $\mathbf{P}_k$. This is accomplished by calculating the difference between the measurement predicted by $\hat{\boldsymbol{x}}_k$ and the actual measurement $\boldsymbol{y}_k$, and then adjusting the prediction to the direction of the difference by an amount that minimizes the mean squared error. [10, 42]

The difference between the measurement and its prediction is called innovation. The measurement model, Eq. (2.51), then leads to the following expressions for the innovation mean $\boldsymbol{v}_k$ and covariance $\mathbf{S}_k$:

$$\boldsymbol{v}_k = \boldsymbol{y}_k - \mathbf{H}_k \hat{\boldsymbol{m}}_k \tag{2.54}$$

$$\mathbf{S}_k = \mathbf{R}_k + \mathbf{H}_k \hat{\mathbf{P}}_k \mathbf{H}_k^\top. \tag{2.55}$$

The minimum mean squared error estimate can then be expressed in terms of the optimal Kalman gain $\mathbf{K}_k$, computed from the innovation covariance by

$$\mathbf{K}_k = \hat{\mathbf{P}}_k \mathbf{H}_k^\top \mathbf{S}_k^{-1}. \tag{2.56}$$

Now the updated posterior distribution mean $\boldsymbol{m}_k$ and covariance $\mathbf{P}_k$ are given by

$$\boldsymbol{m}_k = \hat{\boldsymbol{m}}_k + \mathbf{K}_k \boldsymbol{v}_k \tag{2.57}$$

$$\mathbf{P}_k = \hat{\mathbf{P}}_k - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top, \tag{2.58}$$

which optimally incorporate the information provided by the current measurement into the estimate. [10, 42] The Kalman filter is summarized in Algorithm 2.2.

Algorithm 2.2 The Kalman filter.

1. Choose a prior distribution $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{m}_0, \mathbf{P}_0)$ and set $k = 1$.
2. Recursively calculate $p(\boldsymbol{x}_k | \boldsymbol{y}_{1:k}) = \mathcal{N}(\boldsymbol{m}_k, \mathbf{P}_k)$ from the previous estimate.
   (a) Predict step:

$$\hat{\boldsymbol{m}}_k = \mathbf{A}_{k-1} \boldsymbol{m}_{k-1} \tag{2.59}$$

$$\hat{\mathbf{P}}_k = \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^\top + \mathbf{Q}_{k-1} \tag{2.60}$$

   (b) Update step:

$$\boldsymbol{v}_k = \boldsymbol{y}_k - \mathbf{H}_k \hat{\boldsymbol{m}}_k \tag{2.61}$$

$$\mathbf{S}_k = \mathbf{R}_k + \mathbf{H}_k \hat{\mathbf{P}}_k \mathbf{H}_k^\top \tag{2.62}$$

$$\mathbf{K}_k = \hat{\mathbf{P}}_k \mathbf{H}_k^\top \mathbf{S}_k^{-1} \tag{2.63}$$

$$\boldsymbol{m}_k = \hat{\boldsymbol{m}}_k + \mathbf{K}_k \boldsymbol{v}_k \tag{2.64}$$

$$\mathbf{P}_k = \hat{\mathbf{P}}_k - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top \tag{2.65}$$

3. Set $k = k + 1$ and repeat from step 2 for every new measurement.

2.4.2 Kalman smoother

The smoothing distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:T}) = \mathcal{N}(\boldsymbol{m}_k^{\mathsf{s}}, \mathbf{P}_k^{\mathsf{s}})$ can be computed with backwards recursion from the filtering distributions [42]. This algorithm is known as the Rauch–Tung–Striebel (RTS) smoother [11], or sometimes just as the Kalman smoother. For the final step T the filtering and smoothing distributions naturally coincide, and that serves as the starting point for the backwards recursion. Thus we set $\boldsymbol{m}_T^{\mathsf{s}} = \boldsymbol{m}_T$, $\mathbf{P}_T^{\mathsf{s}} = \mathbf{P}_T$, and consider the step $k < T$. The state evolution model would then predict the following mean and covariance for the succeeding, already calculated step:

$$\hat{\boldsymbol{m}}_{k+1}^{\mathsf{s}} = \mathbf{A}_k \boldsymbol{m}_k \tag{2.66}$$

$$\hat{\mathbf{P}}_{k+1}^{\mathsf{s}} = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^\top + \mathbf{Q}_k. \tag{2.67}$$

Analogously to the Kalman filter update step, this prediction and the already calculated smoothing distribution at the succeeding step are utilized to optimally adjust the current step filtering estimate to recursively include the information from all the succeeding measurements from the filtering distributions. With the help of the gain matrix $\mathbf{G}_k$, the smoothing distribution mean and covariance for the current step can by computed by

$$\mathbf{G}_k = \mathbf{P}_k \mathbf{A}_k^\top \hat{\mathbf{P}}_{k+1}^{-1} \tag{2.68}$$

$$\boldsymbol{m}_k^{\mathsf{s}} = \boldsymbol{m}_k + \mathbf{G}_k(\boldsymbol{m}_{k+1}^{\mathsf{s}} - \hat{\boldsymbol{m}}_{k+1}) \tag{2.69}$$

$$\mathbf{P}_k^{\mathsf{s}} = \mathbf{P}_k + \mathbf{G}_k(\mathbf{P}_{k+1}^{\mathsf{s}} - \hat{\mathbf{P}}_{k+1})\mathbf{G}_k^\top. \tag{2.70}$$

Whereas the filtering distribution can be updated with just a simple recursion step for each measurement, the smoothing distribution requires full backwards recursion calculation whenever a new measurement is obtained. [42] The Kalman smoother is summarized in Algorithm 2.3.

Algorithm 2.3 The Kalman smoother.

1. Calculate the filtering distributions $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) = \mathcal{N}(\boldsymbol{m}_k, \mathbf{P}_k)$ for all $k \leq T$ with the Kalman filter.
2. Set $\boldsymbol{m}_T^{\mathsf{s}} = \boldsymbol{m}_T$, $\mathbf{P}_T^{\mathsf{s}} = \mathbf{P}_T$ and $k = T - 1$.
3. Recursively calculate $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:T}) = \mathcal{N}(\boldsymbol{m}_k^{\mathsf{s}}, \mathbf{P}_k^{\mathsf{s}})$

$$\hat{\boldsymbol{m}}_{k+1}^{\mathsf{s}} = \mathbf{A}_k \boldsymbol{m}_k \tag{2.71}$$

$$\hat{\mathbf{P}}_{k+1}^{\mathsf{s}} = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^\top + \mathbf{Q}_k \tag{2.72}$$

$$\mathbf{G}_k = \mathbf{P}_k \mathbf{A}_k^\top \hat{\mathbf{P}}_{k+1}^{-1} \tag{2.73}$$

$$\boldsymbol{m}_k^{\mathsf{s}} = \boldsymbol{m}_k + \mathbf{G}_k(\boldsymbol{m}_{k+1}^{\mathsf{s}} - \hat{\boldsymbol{m}}_{k+1}) \tag{2.74}$$

$$\mathbf{P}_k^{\mathsf{s}} = \mathbf{P}_k + \mathbf{G}_k(\mathbf{P}_{k+1}^{\mathsf{s}} - \hat{\mathbf{P}}_{k+1})\mathbf{G}_k^\top \tag{2.75}$$

4. Set $k = k - 1$ and repeat from step 3 until $k = 0$.

3 METHODS

This chapter describes methods for the robust estimation of the DFA scaling exponent. The first section proposes a method for determining alpha spectra $\alpha(s)$ as a function of the scale. The method is based on a probabilistic state space model that is solved by Kalman smoother. The latter section presents a versatile optimization scheme for obtaining segmentation of the logarithmic fluctuation function into approximately linear regions. This thesis provides an extended exposition of the subject that was presented at the Computing in Cardiology 2018 conference [18].

3.1 Probabilistic state space model

The alpha spectra are estimated from the logarithmic fluctuation functions $\tilde{F}(\tilde{s})$. The model considers how the scaling exponent varies as a function of the logarithmic window sizes $\tilde{s}$. Therefore, the following notation is adopted. Let $\tilde{s}_k$ denote the utilized logarithmic DFA window sizes in ascending order with $\tilde{F}_k = \tilde{F}(\tilde{s}_k)$ and $\Delta\tilde{F}_k = \Delta\tilde{F}(\tilde{s}_k)$.

The scaling exponent to be estimated, $\alpha(\tilde{s})$, is equivalent to the first derivative of the logarithmic fluctuation function $\tilde{F}(\tilde{s})$ with respect to the logarithmic window sizes $\tilde{s}$. Denoting this differentiation by a prime yields the relation $\alpha(\tilde{s}_k) = \tilde{F}'_k$. The scaling exponent is assumed to remain approximately constant between adjacent window sizes, except for tiny perturbations derived from the data. These perturbations are applied to this derivative $\tilde{F}'_k$. Alternatively, the perturbations may be applied to higher order derivatives, in which case the order of the derivative is also considered the order of the underlaying model. Thus the hidden states $\boldsymbol{x}_k$ of the model for the first and second order models are

$$\boldsymbol{x}_k^{(1)} = \begin{bmatrix} \tilde{F}_k & \tilde{F}'_k & \end{bmatrix}^\top \tag{3.1}$$

$$\boldsymbol{x}_k^{(2)} = \begin{bmatrix} \tilde{F}_k & \tilde{F}'_k & \tilde{F}''_k & \end{bmatrix}^\top, \tag{3.2}$$

where the superscript in parenthesis denotes the order of the model. The measurements $\boldsymbol{y}_k = \tilde{F}_k$ are simply the logarithmic fluctuation function values calculated by DFA. These

lead to the following state transition and measurement model matrices

$$\mathbf{A}_k^{(1)} = \begin{bmatrix} 1 & h_{k_+} \\ 0 & 1 \end{bmatrix} \qquad \mathbf{H}^{(1)} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{3.3}$$

$$\mathbf{A}_k^{(2)} = \begin{bmatrix} 1 & h_{k_+} & \frac{1}{2}h_{k_+}^2 \\ 0 & 1 & h_{k_+} \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{H}^{(2)} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}. \tag{3.4}$$

Here $h_{k_+} = \tilde{s}_{k+1} - \tilde{s}_k$ are the forward differences of the window size at k. The error estimate of the fluctuation function is readily utilized as the measurement noise covariance by

$$R_k = \left[\Delta\tilde{F}(\tilde{s}_k)\right]^2, \tag{3.5}$$

which is now simply a scalar.

Deriving a good estimate for the process noise is the most ambiguous aspect of the model. A particular approach taken here assumes constant magnitude of the noise across the different window sizes. The magnitude of the fluctuations that the perturbed derivative experiences is approximated by a weighted sample variance of a finite difference approximation of the said derivative. The following central difference schemes with non-uniform spacing [43] are considered[1]:

$$\hat{\tilde{F}}'_k \approx \frac{h_{k_-}^2 \tilde{F}_{k+1} + \left(h_{k_+}^2 - h_{k_-}^2\right)\tilde{F}_k - h_{k_+}^2 \tilde{F}_{k-1}}{h_{k_-} h_{k_+} \left(h_{k_+} + h_{k_-}\right)} \tag{3.6}$$

$$\hat{\tilde{F}}''_k \approx \frac{2h_{k_-} \tilde{F}_{k+1} - 2\left(h_{k_+} + h_{k_-}\right)\tilde{F}_k + 2h_{k_+} \tilde{F}_{k-1}}{h_{k_-} h_{k_+} \left(h_{k_+} + h_{k_-}\right)}, \tag{3.7}$$

where $h_{k_-} = \tilde{s}_k - \tilde{s}_{k-1}$ are the backward differences of the logarithmic window size at k. The error arising from the uncertainty in the fluctuation function may then be estimated by

$$\Delta\hat{\tilde{F}}'^2_k \approx \left[\frac{1}{h_{k_-} h_{k_+} \left(h_{k_+} + h_{k_-}\right)}\right]^2 \left[\left(h_{k_-}^2 \Delta\tilde{F}_{k+1}\right)^2 + \left(h_{k_+}^2 - h_{k_-}^2\right)^2 \Delta\tilde{F}_k^2 + \left(h_{k_+}^2 \Delta\tilde{F}_{k-1}\right)^2\right] \tag{3.8}$$

$$\Delta\hat{\tilde{F}}''^2_k \approx \left[\frac{2}{h_{k_-} h_{k_+} \left(h_{k_+} + h_{k_-}\right)}\right]^2 \left[\left(h_{k_-} \Delta\tilde{F}_{k+1}\right)^2 + \left(h_{k_+} + h_{k_-}\right)^2 \Delta\tilde{F}_k^2 + \left(h_{k_+} \Delta\tilde{F}_{k-1}\right)^2\right], \tag{3.9}$$

The boundaries are considered by assuming that the function continues linearly.

[1] Setting $h = h_{k_-} = h_{k_+}$ recovers the standard central difference approximations.

The weighted sample variance $\hat{\sigma}^2$ of the derivatives is computed with the weights taken to be inversely proportional to these squared error estimates. This variance estimates the magnitude of the nudges that the perturbed derivative experiences between logarithmic window sizes, leading to the following process noise covariances:

$$\mathbf{Q}_k^{(1)} = \hat{\sigma}^2 \begin{bmatrix} \frac{1}{3}h_{k_+}^3 & \frac{1}{2}h_{k_+}^2 \\ \frac{1}{2}h_{k_+}^2 & h_{k_+} \end{bmatrix} \tag{3.10}$$

$$\mathbf{Q}_k^{(2)} = \hat{\sigma}^2 \begin{bmatrix} \frac{1}{20}h_{k_+}^5 & \frac{1}{8}h_{k_+}^4 & \frac{1}{6}h_{k_+}^3 \\ \frac{1}{8}h_{k_+}^4 & \frac{1}{3}h_{k_+}^3 & \frac{1}{2}h_{k_+}^2 \\ \frac{1}{6}h_{k_+}^3 & \frac{1}{2}h_{k_+}^2 & h_{k_+} \end{bmatrix}. \tag{3.11}$$

Finally an estimate is required for the prior distribution $\mathcal{N}(\boldsymbol{m}_0, \mathbf{P}_0)$. This may simply be estimated from the first data points:

$$\boldsymbol{m}_0^{(1)} = \begin{bmatrix} \tilde{F}_1 \\ \hat{\tilde{F}}_1' \end{bmatrix} \qquad \mathbf{P}_0^{(1)} = \begin{bmatrix} \Delta\tilde{F}_1^2 & 0 \\ 0 & \Delta\hat{\tilde{F}}_1'^2 \end{bmatrix} \tag{3.12}$$

$$\boldsymbol{m}_0^{(2)} = \begin{bmatrix} \tilde{F}_1 \\ \hat{\tilde{F}}_1' \\ \hat{\tilde{F}}_1'' \end{bmatrix} \qquad \mathbf{P}_0^{(2)} = \begin{bmatrix} \Delta\tilde{F}_1^2 & 0 & 0 \\ 0 & \Delta\hat{\tilde{F}}_1'^2 & 0 \\ 0 & 0 & \Delta\hat{\tilde{F}}_1''^2 \end{bmatrix}. \tag{3.13}$$

The Kalman filter and smoother may then be applied to the specified model to obtain the filtering and smoothing posterior distributions $\mathcal{N}(\boldsymbol{m}_k, \mathbf{P}_k)$ and $\mathcal{N}(\boldsymbol{m}_k^{\mathsf{s}}, \mathbf{P}_k^{\mathsf{s}})$.

3.2 Linear programming model

Let us consider the problem of optimally segmenting the logarithmic fluctuation function $\tilde{F}_k$ into approximately linear regimes. The problem is considered in two parts. First the optimal segmentation is solved for fixed numbers of segments N. The optimal value for this number is then sought from these solutions.

The algorithm proceeds by computing all the possible linear regression fits to the logarithmic fluctuation function. If desired, additional requirements could be enforced for the fitted lines, such as minimum length. Let n be the number of linear fits and RSS_i be the residual sum of squares of the i-th fit. The optimal segmentation minimizes the total residual sum of squares. The optimization problem is approached from integer linear programming[2] (ILP) viewpoint. In linear programming the objective function and constraints are linear, and furthermore in ILP the variables are constrained to be integers. [44]

[2]Linear optimization is called linear programming for historical reasons.

The ILP problem may be stated as follows:

$$\arg\min_{\boldsymbol{x}} \boldsymbol{c}^\top \boldsymbol{x} \qquad \boldsymbol{c} \in \mathbb{R}^n \qquad \boldsymbol{x} \in \{0,1\}^n, \tag{3.14}$$

where the components x_i of the binary vector $\boldsymbol{x}$ indicate whether the i-th fit is utilized in the segmentation. The components c_i of the coefficient vector $\boldsymbol{c}$ are the residual sum of squares RSS_i of the linear regression fits. The optimization constraints are expressed with the help of auxiliary variables a_{ik}, defined to be equal to unity if the i-th fit covers the k-th window size and zero otherwise. The constraints are then given by the following equations

$$\sum_{i=1}^{n} a_{ik} = 1; \qquad \sum_{i=1}^{n} x_i = N, \tag{3.15}$$

where the former relations ensure that each point in the fluctuation function is covered by exactly one linear fit, and the latter relation ensures that the solution consists of the desired number of linear fits N. The problem is then readily solved by any existing integer linear programming suite, such as the open source GNU Linear Programming Kit (GLPK).

Choosing an optimal value for the number of linear segments, N, is a partial and subjective decision. The approach taken here proceeds as follows. Let $\mathrm{RSS}(N)$ denote the total residual sum of squares of the optimal segmentation with N segments. This residual is trivially reduced by increasing the number of segments. We seek a solution that consists of as few segments as admissible for equitable segmentation. Therefore, we consider the desirability $D(N)$ of a solution to be inversely proportional to the number of segments and the total residual sum of squares:

$$D(N) \propto \frac{1}{N \cdot \mathrm{RSS}(N)}. \tag{3.16}$$

The solution is considered optimal when it maximizes this quantity. Graphical illustration of this algorithm is presented in Fig. 3.1.

This optimization scheme may be performed separately for each fluctuation function or simultaneously to a group of fluctuation functions, provided that they all have been calculated at compatible window sizes. In the latter case the residual sum of squares RSS_{ij} of the i-th fit in the j-th fluctuation function are summed over all the fluctuation functions when computing the coefficients as $c_i = \sum_j \mathrm{RSS}_{ij}$.

An example of these methods for the alpha spectra and optimal segmentation is shown in Fig. 3.2, where the methods are applied to real data. By visual inspection it can be noticed that the resulting linear segmentation is sensible and the estimated alpha spectra smoothly follows the noisy fluctuation function.

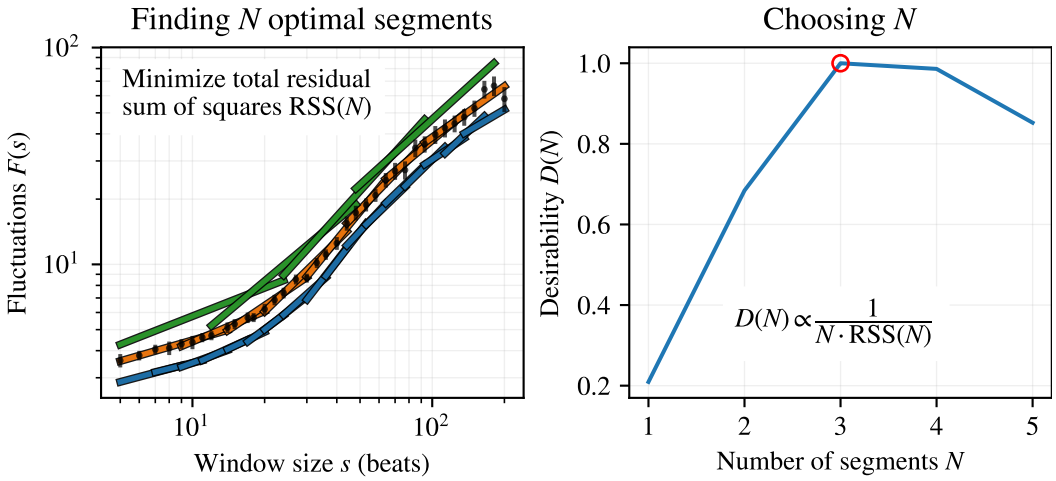

Figure 3.1. Illustration of the optimal segmentation algorithm. *First all the possible linear regression fits to the logarithmic fluctuation function are computed at all possible scales. For illustrative purposes only some of the fits are depicted. The solution with N fits and the lowest total residual sum of squares RSS(N) is obtained by integer linear programming. The number of segments N is chosen by maximizing the quantity $1/\left[N \cdot RSS(N)\right]$. The input time series for computing the fluctuation function by DFA1 is a snippet of RR intervals from the database of Ref. [45]*

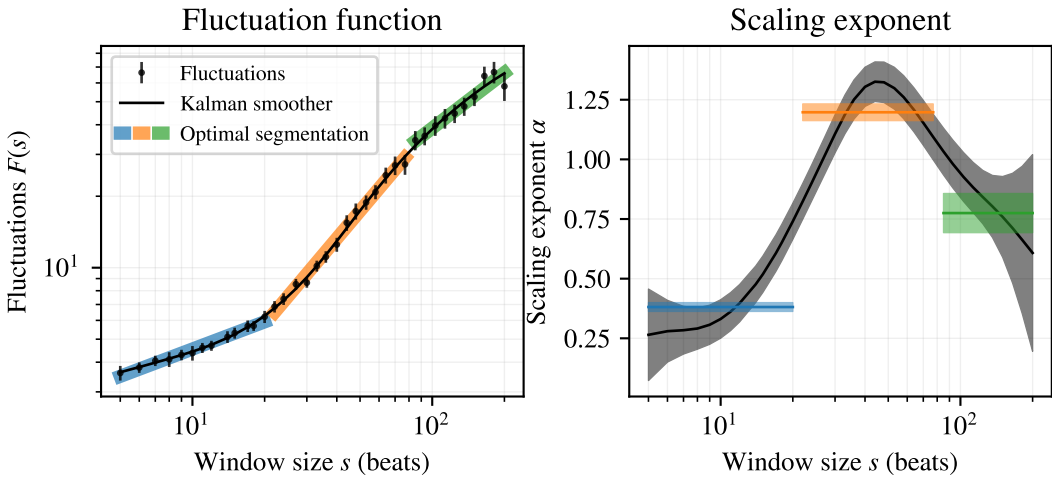

Figure 3.2. Example of the new methods for determining the scaling exponents. *The 1st order model is utilized for computing the alpha spectra by Kalman smoother. The error bounds for the alpha spectra are calculated from the posterior distribution covariance matrix. The error bounds for the linear segmentation denote the standard error of the linear regression estimate. The input time series for computing the fluctuation function by DFA1 is a snippet of RR intervals from the database of Ref. [45].*

4 APPLICATIONS AND RESULTS

4.1 Validation of the methods

4.1.1 Artificial test data

Consistent evaluation of the methods requires a reliable algorithm for generating artificial time series with known scaling exponents. The simulation of $1/f^\beta$ noises with arbitrary exponents β is straightforward by the so-called Fourier filtering algorithm. The technique proceeds by generating uncorrelated normally distributed random numbers, and then enforcing the desired power law scaling in the Fourier frequency space [46]. The procedure is described in Algorithm 4.1.

Time series generated in such a manner may be considered to be pure $1/f$ noises by construction, but they are only approximations to fractional Gaussian noises and Brownian motions. The method also suffers from inaccuracies arising from periodic effects and singularity in the autocovariance function [47]. Several methods exist for the exact generation of fGn and fBm, and the algorithm considered here was introduced by Davies and Harte [48]. The method, presented in Algorithm 4.2, also relies on Fourier transforms for the computation of eigenvalues of a circulant matrix, resulting in one of the most efficient algorithms for their generation, as the Fast Fourier Transform can be exploited [48, 49].

Algorithm 4.1 The Fourier filtering algorithm for generating $1/f^\beta$ noises.

1. Generate N independent normally distributed random numbers u_i for $1 \leq i \leq N$.
2. Compute the Fourier transform of u_i to obtain the Fourier transform coefficients c_i and the corresponding frequencies f_i.
3. Compute scaled coefficients by $c_i' = |f_i|^{-\beta/2} c_i$ to enforce the desired scaling.
4. Perform inverse Fourier transform of the scaled coefficients c_i' to obtain a sample of $1/f^\beta$ noise.

4.1.2 Numerical results

As was discussed in Sec. 2.3.2, the detrending in DFA introduces bias to the fluctuation function and consequently to the estimated scaling exponents. In case of fGn and fBm the analytical autocovariance function, Eq. (2.7), and variogram, Eq. (2.6), are known. This permits the computation of the theoretical expected DFA fluctuation function for these processes, as described in Sec. 2.3.1. The scaling exponent is then obtained by differentiating the logarithmic fluctuation function with respect to the logarithmic window sizes.

Algorithm 4.2 The Davies-Harte algorithm for generating fGn and fBm.

1. To generate a sample of length N, compute the autocovariances $C_H(n)$ of fGn with the desired Hurst parameter H by Eq. (2.7) for $0 \leq n \leq N$ and set[1]

$$\alpha_k = \begin{cases} C_H(k) & 0 \leq k \leq N \\ C_H(2N-k) & N+1 \leq k \leq 2N-1 \end{cases} \tag{4.1}$$

   for $0 \leq k \leq 2N-1$.
2. Compute the Fourier transform of α_k and denote it by λ_k. These λ_k correspond to the eigenvalues of the circulant matrix whose first row is α_k. These eigenvalues are real and non-negative by construction.
3. Generate independent normally distributed random numbers $U_k^{\Re}$ and $U_k^{\Im}$ to compute

$$w_k = \begin{cases} \sqrt{\frac{\lambda_k}{2N}} U_k^{\Re} & k = 0, N \\ \sqrt{\frac{\lambda_k}{4N}} \left(U_k^{\Re} + i U_k^{\Im} \right) & 1 \leq k \leq N-1 \\ \sqrt{\frac{\lambda_k}{4N}} \left(U_{2N-k}^{\Re} - i U_{2N-k}^{\Im} \right) & N+1 \leq k \leq 2N-1 \end{cases} \tag{4.2}$$

   for $0 \leq k \leq 2N-1$.
4. Compute the Fourier transform of w_k and denote it by z_k.
5. (a) $x_k = z_k$ is a sample of fGn for $0 \leq k \leq N-1$.
   (b) $x_k = \sum_{i=0}^{k} z_i$ is a sample of fBm for $0 \leq k \leq N-1$.

The derivation is carried out by the finite difference scheme of Eq. (3.6) to avoid cumbersome analytical differentiation. As these functions behave smoothly, this numerical scheme is justified and expected to introduce negligible error.

These results are illustrated in Fig. 4.1(a). For ease of visualization the results are plotted as deviation from the expected asymptotic scaling. This asymptotic scaling exponent is equal to the Hurst parameter for fGn, and for fBm it is equal to the Hurst parameter plus one. The scaling exponent is overestimated in small window sizes for fGn ($0 < \alpha < 1$). While the behavior of the deviation stays approximately constant for persistent noise ($1/2 < \alpha < 1$), the deviation begins to extend to larger window sizes at an accelerating pace as the noise gets more and more antipersistent ($0 < \alpha < 1/2$). The author is not aware of other work that studies the theoretical scaling exponent as a function of the window size, but these findings are in agreement with the consensus that there exists a bias at the small scales and that DFA becomes less accurate with strongly anticorrelated data.

The behavior becomes more peculiar when transitioning into the fBm regime ($1 < \alpha < 2$). Motions with anticorrelated increments ($1 < \alpha < 1 1/2$) exhibit a prominent dip in the scaling exponent that extends to scales over 1000 for Hurst parameters close to zero. This is a striking result that has consequences for reliable classification of processes with scaling exponents close to unity. A single scaling exponent obtained from a simple linear regression is particularly prone to misclassification in this instance. On the other hand,

[1] Some sources [49] mention setting $\alpha_N = 0$ instead of $C_H(N)$. However, this may cause the algorithm to fail for H close to 1 with short sample lengths, as some of the eigenvalues in step 2 may be negative.

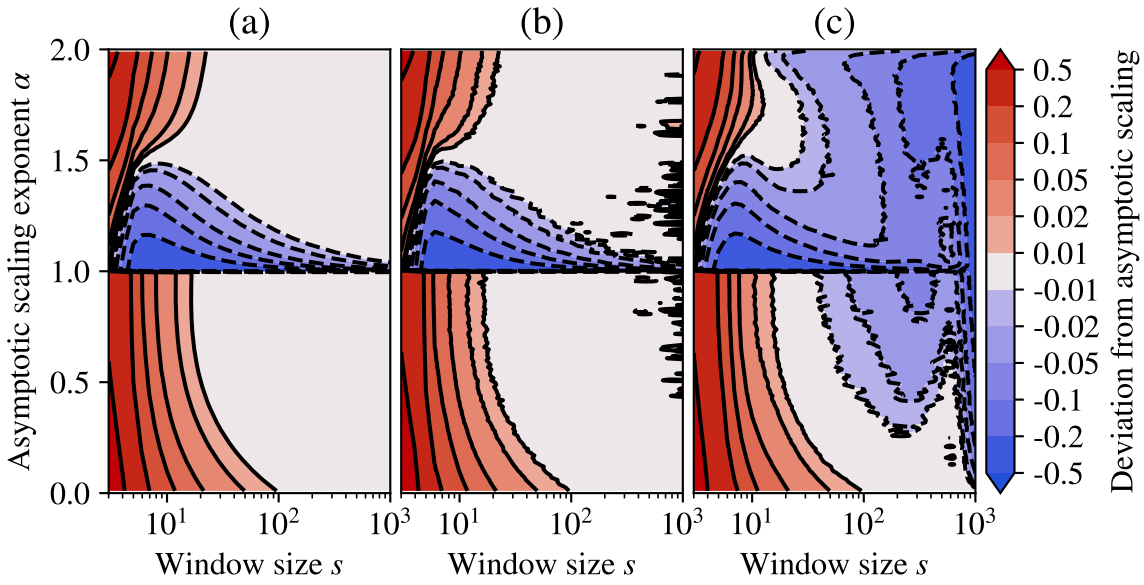

Figure 4.1. The deviation of the DFA scaling exponent from asymptotic behavior for fractional Gaussian noise (fGn) and fractional Brownian motion (fBm). *Fractional Gaussian noises have asymptotic DFA scaling exponents* $0 < \alpha < 1$ *equal to the Hurst parameter* H*. Fractional Brownian motions have asymptotic DFA scaling exponents* $1 < \alpha < 2$ *with the relationship to the Hurst parameter given by* $\alpha = H + 1$*. In (a) the theoretical DFA fluctuation functions have been computed by Eq.* (2.30) *and Eq.* (2.31)*, utilizing the exact autocovariances and variograms of fGn and fBm, respectively. The scaling exponent is computed by finite difference approximation from the fluctuation function, yielding theoretical deviation from the asymptotic scaling behavior. Simulated results in (b) and (c) are obtained by computing 10 000 realizations of fGn and fBm by the Davies-Harte method for each value of the Hurst parameter. The scaling exponents have been calculated by the Kalman smoother method with 2nd order state space model. In (b) only a single fluctuation function has been computed from the 10 000 realizations that includes the squared fluctuations of Eq.* (2.15) *from all the realizations. In (c) the fluctuation functions are separately computed for each realization, and the scaling exponent is the average from all the fluctuation functions. This averaging process is observed to introduce bias particularly at the larger window sizes, as the significance of rarer larger fluctuations is diminished.*

utilizing the full alpha spectra may allow more informed decision to be made, as the details and shape of the curve will differ for fGn and fBm with the scaling exponent close to unity. For persistent increments ($1^1/_2 < \alpha < 2$) the scaling exponent is again overestimated at the smallest scales, with the effect becoming stronger with increasing Hurst parameter. Curiously, the pure Brownian motion ($\alpha = 1^1/_2$) is the least biased member in the family of fGn/fBm processes.

The validity of the alpha spectra obtained by the Kalman smoother is verified by comparing the spectra obtained from simulated data to these theoretically predicted values. Instances of fGn and fBm are simulated by the Davies-Harte method, which produces samples with exact covariance structure for these processes. For each value of the Hurst parameter 10 000 samples of length $N = 1000$ are generated to obtain sufficient statistical accuracy. Two different methods are considered when averaging over the samples. In

the first approach only a single fluctuation function is computed that includes the squared fluctuations of Eq. (2.15) from all the realizations. Alternatively fluctuation functions are calculated for each sample, from which the alpha spectra are estimated and then averaged. As discussed in Sec. 2.2, this latter approach is mathematically wrong, but may yet still be the preferred solution when studying actual data in practice.

The results with single fluctuation function for each sample are shown in Fig. 4.1(b), and these are in an excellent concordance with the theoretical prediction. Slight noise is visible in the results, particularly at the very largest scales with the scaling exponent close to unity. This is most likely a finite size statistical effect, and no systematic bias is observed that would arise from the Kalman smoother method. All of the simulated results utilize the 2nd order state space model in the Kalman smoother method, as it was found to produce smoother estimates.

As is observed in Fig. 4.1(c), and anticipated by theoretical arguments, averaging over the fluctuation functions introduces bias. The scaling exponent is underestimated, relative to the expected value, as the significance of larger fluctuations is diminished. While the bias is evident, its magnitude does not exceed 0.1 for scales less than the generally recommended maximum value $N/4$.

The Davies-Harte method produces exact samples of fGn and fBm, which in part is also experimentally verified by the results of Fig. 4.1(b). It has been shown that these processes exhibit asymptotic scaling in the frequency domain that corresponds to $1/f$ noises. However, it is not clear how exact scaling in the frequency domain, as enforced by the Fourier filtering method, corresponds to the DFA scaling exponent as the function of the DFA scale. This question is of practical interest as in the literature Fourier filtering is possibly the most popular method for generating artificial test data for DFA. Sometimes even conclusions are drawn with the assumption that these Fourier filtering samples represent fGn and fBm [6, 19, 27, 40]. The differences in the scaling behavior for these time series generation methods are illustrated in Fig. 4.2.

It is immediately clear that the $1/f$ noises generated by Fourier filtering in Fig. 4.2(b) do not correspond to exact fGn and fBm. The periodicity of the Fourier transform introduces bias at the largest scales, and the correct asymptotic scaling is only reached for pure white noise. The effect of the periodicity could be compensated by generating longer samples and discarding elements from the beginning and the end. This is only mildly effective, and even a 1000-fold increase in length does not restore the expected asymptotic scaling. A possible reason for the too small scaling exponents in the $\alpha > 1$ regime could be due to the fact that the autocovariance function obtained as the Fourier transform of the power spectral density is always stationary by construction. The small increase in the bias at the very largest scales in the anticorrelated regime does get eliminated by the longer sampling. The very prominent bias across all the scales for anticorrelated samples explains the claims that DFA is not very accurate for highly anticorrelated time series, as many of these studies utilized Fourier filtering for their synthetic data [50].

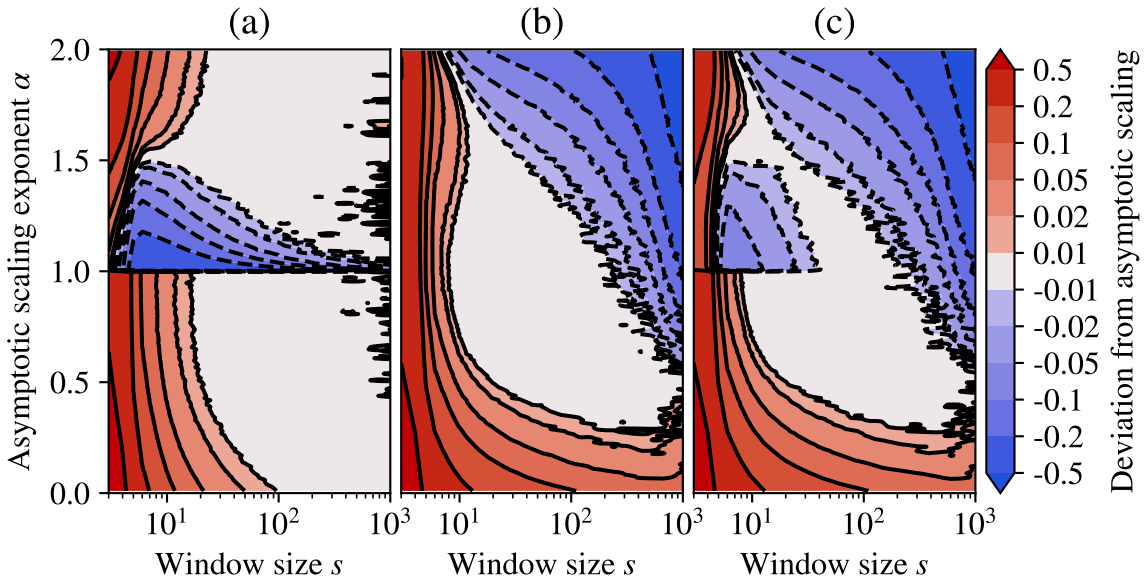

Figure 4.2. The deviation of the DFA scaling exponent from asymptotic behavior for different time series generation methods. *Exact fGn and fBm by the Davies-Harte method (a), long-range correlated data by Fourier filtering (b), and direct Fourier filtering for the increment process with* $0 < \alpha \leq 1$ *and cumulative sum of this increment process for* $1 < \alpha \leq 2$ *(c). In each case 10 000 realizations are simulated for each value of the asymptotic scaling exponent* α*. The scaling exponents are calculated as explained in the caption for Fig. 4.1(b).*

Discrete fBm is constructed as the cumulative sum of fGn. Similar approach for $1/f$ noises of Fourier filtering is considered in Fig. 4.2(c), where direct Fourier filtering is utilized for the increment process with $0 \leq \alpha < 1$, and the cumulative sum of these increments for $1 \leq \alpha < 2$. It would appear that the dip in the scaling exponent at the small scales of non-stationary processes with antipersistent increments is a side effect of the cumulative summation. This raises further questions how transforming the input time series by taking differences or by repeated cumulative summation affects the details of the scaling, particularly at the smallest scales. The scaling at the larger window sizes remains consistent, regardless whether the time series were constructed by cumulative summation.

For practical purposes it is not only important that the mean scaling exponents converge to the correct values, but also their variance should be minimal across different realizations of the same stochastic process. The standard deviations of the scaling exponents are shown in Fig. 4.3. The standard deviation is estimated by simulating 10 000 samples for each value of the asymptotic scaling exponent with the three discussed methods for generating the time series. The fluctuation function is computed separately for each sample, in order to be able to obtain alpha spectra for each sample. The alpha spectra are estimated by the Kalman smoother as before, from which the standard deviation is calculated for each window size. There is no appreciable variation amongst the different time series generation methods. As expected, the standard deviation increases with increasing window sizes, as the number of available non-overlapping windows is reduced. The standard deviation decreases towards the extremal detectable DFA1 scaling exponents

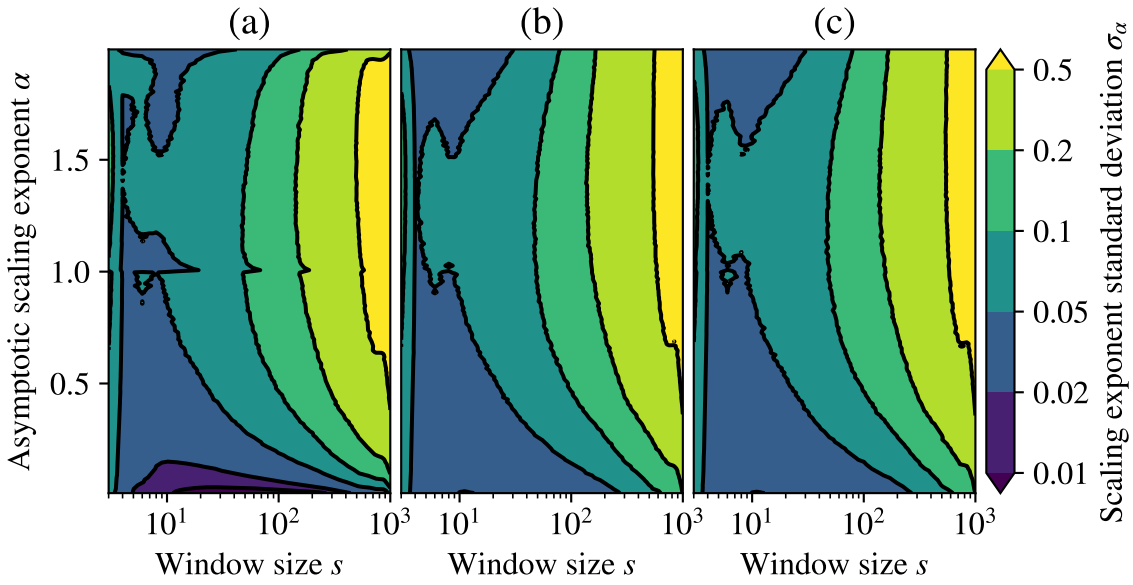

Figure 4.3. The standard deviations of the scaling exponents with different time series generation methods. *Exact fGn and fBm by the Davies-Harte method (a), long-range correlated data by Fourier filtering (b), and direct Fourier filtering for the increment process with* $0 < \alpha \leq 1$ *and the cumulative sum of this increment process for* $1 < \alpha \leq 2$ *(c). In each case 10 000 realizations are simulated for each value of the asymptotic scaling exponent* α*. The scaling exponents are calculated as explained in the caption for Fig. 4.1(c).*

0 and 2, which is understandable, as at the extremes there is less room for variation. There is less deviation for stationary processes, particularly for the most antipersistent fGn, when compared to the non-stationary processes. Maximal standard deviation is found in non-stationary processes with antipersistent increments.

The effect of the order of DFA detrending on the variance is studied in Fig. 4.4. The variance is found to increase with increasing order of DFA. Keep in mind that the maximum scaling exponent that can be detected by DFAn is $n + 1$, and larger exponents will saturate at this value. This explains the low variance regions for the larger exponents for DFA0 and DFA1. The effect of DFA order is further studied in Figs. 4.5 through 4.7 for the different time series generation methods, together with the consequences of the biased computation of the averages. The results are compatible with the previous knowledge that increasing the order of the DFA increases the bias in the fluctuation function. However, the behavior near the maximal detectable exponent complicates this conclusion. Fourier filtered samples appear to better approach asymptotic scaling with increasing order of DFA. The bias introduced by improper averaging procedure can be dangerous for interpreting results of numerical experiments, if it is not properly acknowledged.

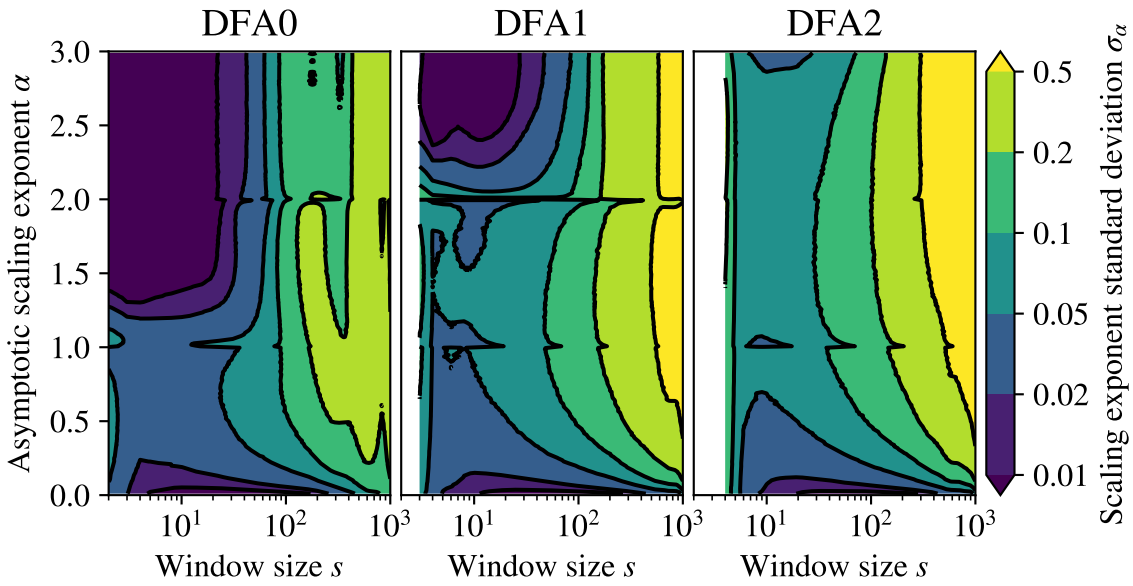

Figure 4.4. The standard deviations of the scaling exponents with different orders of DFA. *For each value of the asymptotic scaling exponent 10 000 realizations of fGn are simulated by the Davies-Harte method. For* $0 < \alpha < 1$ *the fGn with Hurst parameter* $H = \alpha$ *are utilized directly. For* $1 < \alpha < 2$ *cumulative sums of fGn with* $H = \alpha - 1$ *are utilized, resulting in fBm. For* $2 < \alpha < 3$ *the cumulative summation is carried out twice, corresponding to integrated fBm with* $H = \alpha - 2$*. The scaling exponents are calculated as explained in the caption for Fig. 4.1(c).*

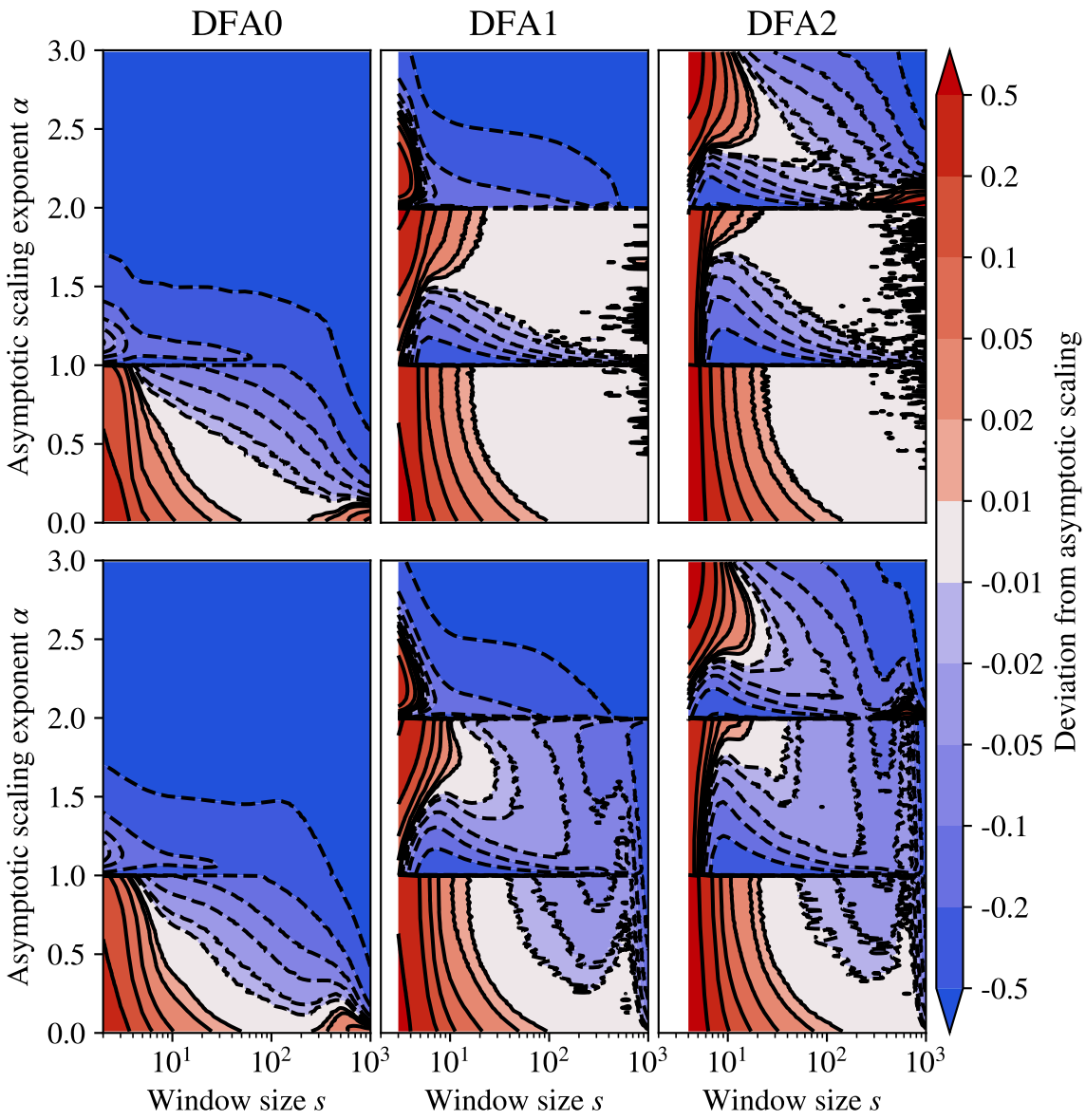

Figure 4.5. Comparison of DFA results for fGn as the increment process with different orders of DFA. *In the upper panels only single fluctuation functions have been computed for each value of α, as described in the caption for Fig. 4.1(b). In the lower panels average scaling exponents have been calculated as described in the caption for Fig. 4.1(c).*

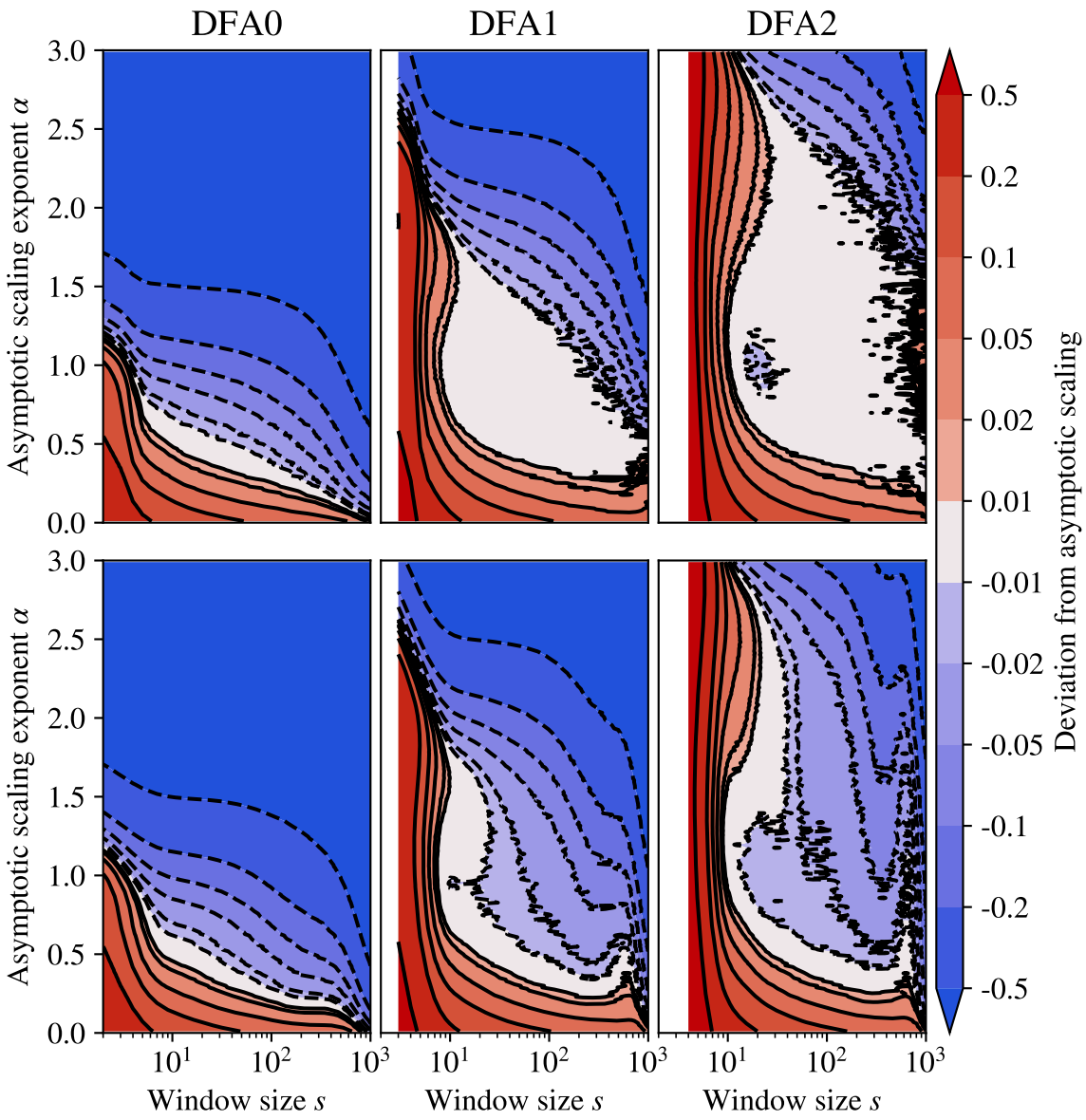

Figure 4.6. Comparison of DFA results for Fourier filtering with different orders of DFA. *In the upper panels only single fluctuation functions have been computed for each value of α, as described in the caption for Fig. 4.1(b). In the lower panels average scaling exponents have been calculated as described in the caption for Fig. 4.1(c).*

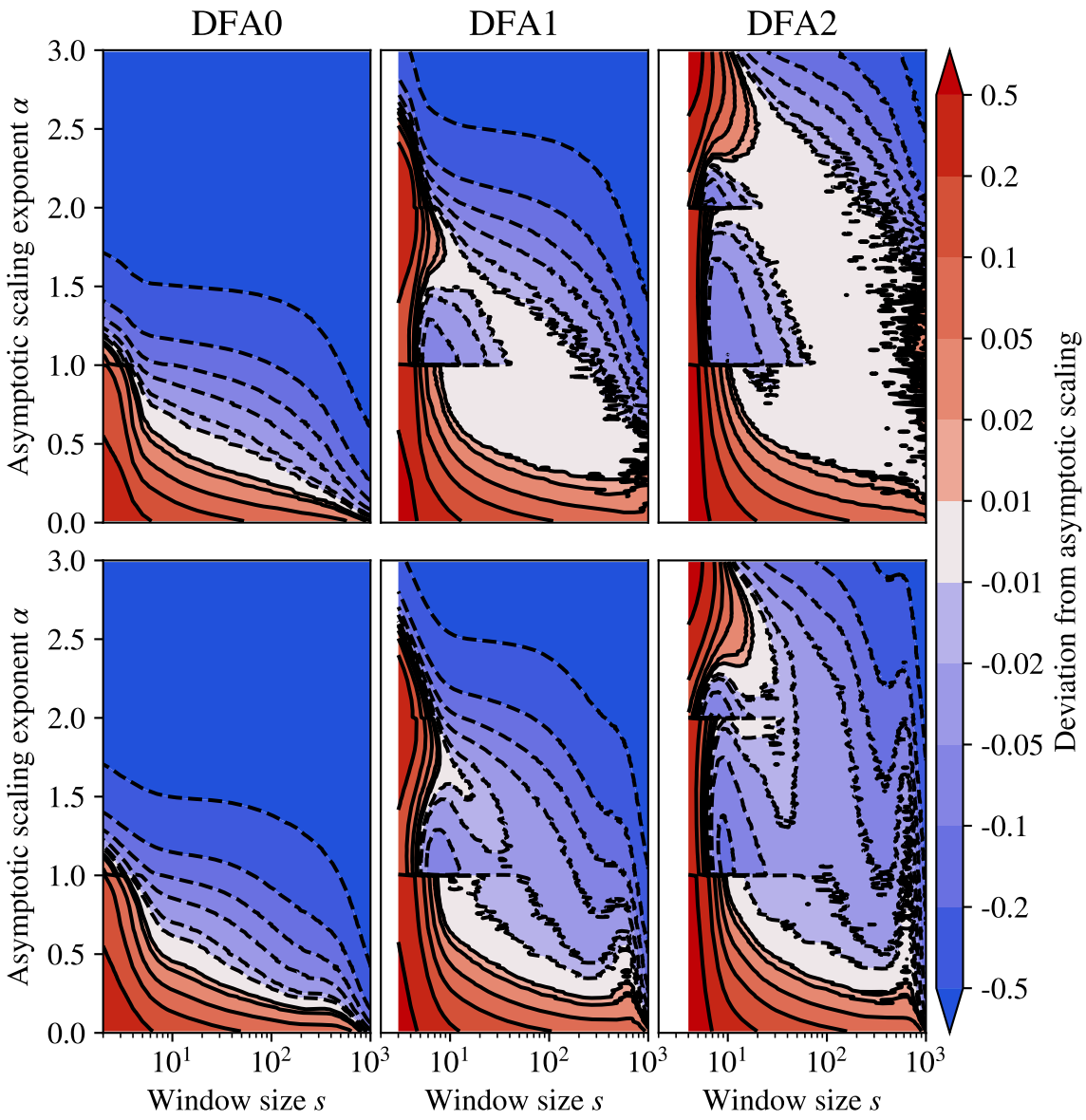

Figure 4.7. Comparison of DFA results for Fourier filtering and cumulative summation with different orders of DFA. *In the upper panels only single fluctuation functions have been computed for each value of α, as described in the caption for Fig. 4.1(b). In the lower panels average scaling exponents have been calculated as described in the caption for Fig. 4.1(c).*

4.2 Heart rate variability

The beating of the heart is regulated by the autonomic nervous system. Heart beats are not precisely regular, and complex variations are characteristic of a healthy heart. Alterations in this heart rate variability (HRV) may be indicative of various cardiovascular diseases and reflect the state of the autonomic nervous system. [3, 51] Intense physical exercise has also been found to profoundly influence HRV [52].

HRV is studied by analyzing the intervals between subsequent heart beats. These intervals are most accurately obtained from the electrocardiogram (ECG), which is a recording of the electrical activity in the heart as a function of time. The components present in a typical ECG are described in Fig. 4.8. The locations of the beats are usually determined from the R peaks of the QRS complexes, which are the most prominent features of the ECG and correspond to the depolarization of the ventricles. [51, 53] For the purposes of this thesis, the interbeat intervals (IBI) are defined as the RR intervals between adjacent R peaks. Their variability may be quantified by several standardized measures in both the time- and frequency-domain. Some examples include the standard deviation of all the intervals, or the power in specific frequency ranges, respectively. [51] However, more sophisticated tools of time series analysis must be utilized to fully appreciate the complex fractal-like variations in the beat rate. Straightforward application of the tools is complicated by non-stationarities present in the heart beat recordings. These non-stationarities give rise to trends that can be accounted for, at least to some extent, by detrended fluctuation analysis, which has become a popular tool for analyzing HRV [2, 3, 13–15, 18, 54].

Conventionally the analysis has been performed by extracting two scaling exponents from linear fits, for short- (α_1) and longe-range (α_2) correlations respectively [2]. However, the extent of these ranges is subjective and the linear two-range model potentially disregards additional information present in the data. This approach has previously been criticized as an over-simplification, and alternative methods have been proposed [13–15, 18, 54]. The tools presented in this thesis are utilized for analyzing the fractal scaling of RR intervals in different cardiac conditions.

4.2.1 Data and methods

The effect of various heart conditions on the fractal scaling behavior of RR intervals is studied by utilizing the publicly available databases of PhysioBank [56]. To establish a reference baseline, the fractal scaling behavior is studied in healthy subjects. The following pathological conditions are investigated: congestive heart failure (CHF), atrial fibrillation (AF), and episodic ST segment variation (ST ep.). Examples of RR intervals from patients with different cardiac conditions are shown in Fig. 4.9, where subtle differences in their correlations may be observed. All of the examples appear to display at least approximate self-similarity across the different scales. The CHF case exhibits higher short-term vari-

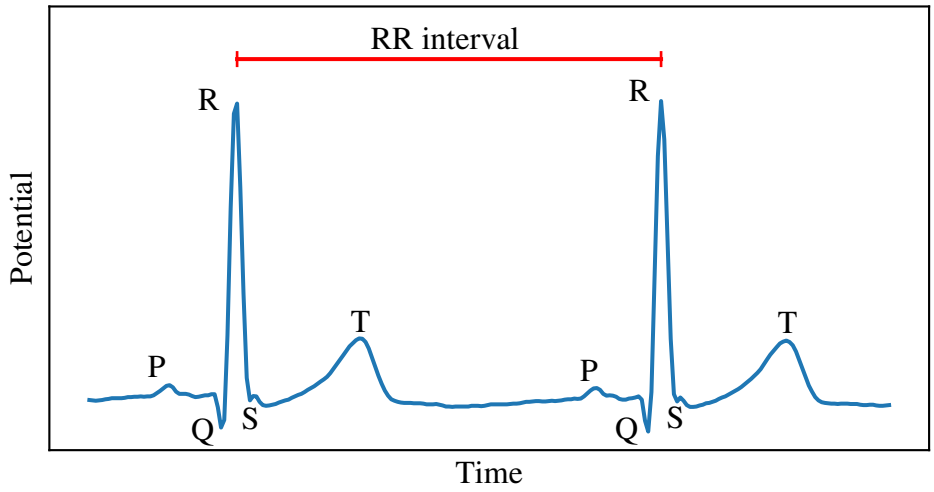

Figure 4.8. Illustration of electrocardiogram. *The P wave is associated with the depolarization of the atria. The QRS complex represents ventricular depolarization, and the T wave results from the repolarization of the ventricles. [53] Based on real ECG from Ref. [55].*

ability and anticorrelations than the healthy example. The overall variability appears to be suppressed in the AF example. However, detailed correlations are next to impossible discern by the eye, and therefore statistical tools are necessary for accurate quantitative assessment.

The following databases of long-term ECG recordings are utilized in the study: "The MIT-BIH Normal Sinus Rhythm Database" (nsrdb [55]), "Normal Sinus Rhythm RR Interval Database" (nsr2db [57]), "The BIDMC Congestive Heart Failure Database" (chfdb [58, 59]), "Congestive Heart Failure RR Interval Database" (chf2db [45]), "The Long-Term AF Database" (ltafdb [60, 61]), and "The Long-Term ST Database" (ltstdb [62, 63]). These databases are summarized in Table 4.1.

Table 4.1. *PhysioBank databases utilized in the thesis.*

Database	Condition	Number of subjects	Sampling frequency (Hz)
nsrdb	Healthy	18	128
nsr2db	Healthy	54	128
chfdb	CHF	15	250
chf2db	CHF	29	128
ltafdb	AF	84	128
ltstdb	ST ep.	86	250

The already annotated beats of the aforementioned databases are utilized for the analy-

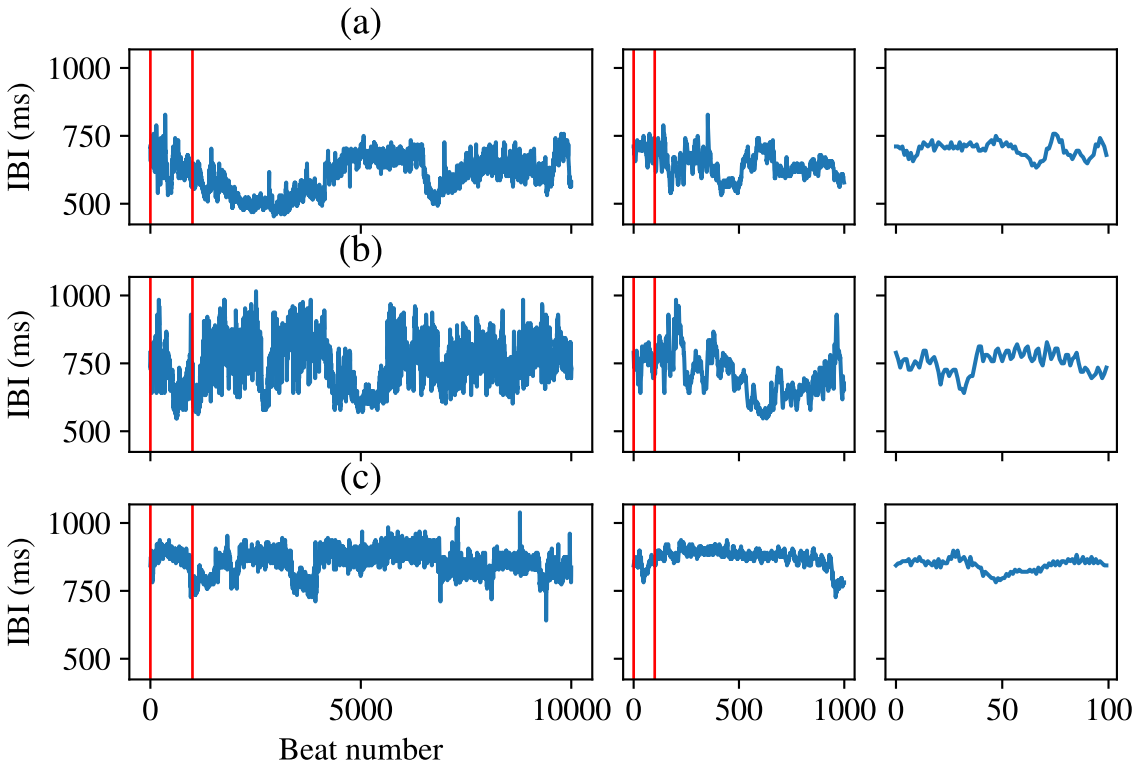

Figure 4.9. Illustration of heart rate variability and fractal scaling. *Shown are examples of RR interval recordings from the PhysioBank[56] data at multiple scales for healthy (a), congestive heart failure (b), and atrial fibrillation (c).*

sis, removing the necessity of detecting the beats from ECGs. The annotations consist of the time and type of the beat. Therefore the RR intervals are obtained as the difference in time between consecutive annotated beats. This raw time series may require preprocessing before the analysis to remove artifacts. The most common type of artifact arises when beats or entire segments of the ECG are missing, resulting in erroneously long intervals. Other, more subtle, errors may also be present in the data due to spurious detection of the QRS complexes. These are filtered by enforcing minimum and maximum limits for the intervals and their variation between adjacent beats. The intervals are rejected if they fall outside 200–1500 ms range, or if the difference between adjacent intervals is greater than 333 ms.

The DFA fluctuation functions are computed in 100 logarithmically distributed window sizes of 3–30 000 beats. The effect of non-stationarities is studied by performing DFA with detrending orders 0, 1, 2 and 3. The detrending procedure in conventional DFA is limited to local polynomial fits that disregard the global behavior of the time series. The ability of DFA to cope with complex non-linear trends is also disputed [29]. Therefore explicit detrending of the time series prior to the DFA procedure is also experimented with. The trend is determined with a moving median filter with kernel sizes of 51, 101 and 501 beats.

The scaling exponent α is determined from the fluctuation functions with three methods: the traditional linear fits in short (4–16 beats) and long (16–64 beats) range regimes, the

optimal linear segmentation applied to the whole dataset at once, and the alpha spectra obtained with the Kalman smoother. The optimal segmentation is performed for both the whole range of 3–30 000 beats and in the 3–64 beat range for better comparison with the traditional two-range approach. The minimum window size is 3 beats, as it is the minimum number of points for which the linear detrending does not trivially remove all variation. This also enhances the view into the bias at the smallest scales. These scaling exponents are then utilized as features in the classification task of identifying the different cardiac conditions. The classification is performed by various algorithms available in the scikit-learn Python module [64]. Hyperparameters are optimized with a simple grid search.

4.2.2 Results on heart rate variability

The scaling exponents obtained by the traditional segmentation into short- (4–16 beats) and long-range (16–64 beats) regimes are shown in Fig. 4.10(a). Shown are the results obtained by the direct application of DFA1. The explicit detrending by the moving median filter has minor impact on the results with 51 beats wide kernel. For larger kernel sizes the effect is minimal. The scaling behavior is very similar for healthy individuals and for those suffering from ST episodes, and distinguishing these from each other would be next to impossible by merely analyzing the fractal scaling of RR intervals. On the other hand, congestive heart failure and atrial fibrillation show distinctive scaling, both from each other and from the healthy and ST episode classes. This is in agreement with previous research about the subject [3].

The optimal linear segmentation is performed simultaneously for all the fluctuation functions to obtain consistent feature vectors for the classification. The results in the 3–64 beat range are shown in Fig. 4.10(b) for comparison with the two-range model. The first segment consists of only 2 beats due to steep increase in the scaling exponents because of the bias in the fluctuation function at the smallest windows. Curiously the distribution for AF becomes quite concentrated in this segment, albeit having long tails. This could be helpful for classification, which is later confirmed by the results in Fig. 4.15(a). Also CHF appears to be more distinct from the healthy and ST episode cases in the first two segments, which is affirmed by the classification results. In the longer scale segments the AF scaling exponents remain at approximately 0.8, while all the other conditions appear to converge to values slightly above 1.0. It is not immediately clear why the 5th segment appears in the optimal segmentation. While the medians are approximately equal within the last two segments, the distributions are thinner in the final segment, which may be enough to embolden the algorithm for splitting the segments. The results suggest that the shortest scale is the most important for cardiac classification. This may, at least in part, be understood by extrinsic trends in the heart rate masking the intrinsic scaling to some extent at longer scales.

The segmentation is also performed for fluctuation functions calculated on the extended

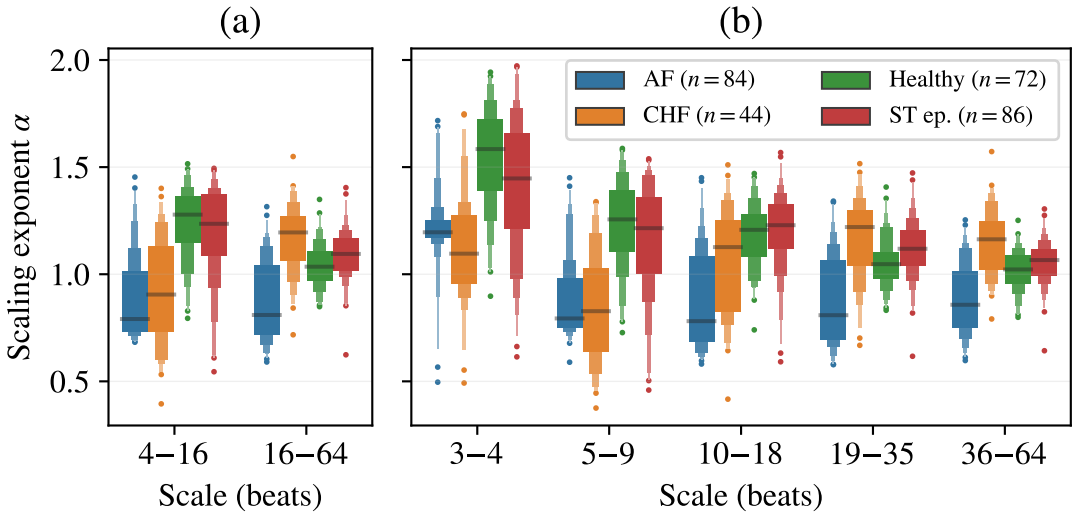

Figure 4.10. Scaling exponents of different cardiac conditions in the 3–64 beat range. *The distributions are illustrated by so-called letter-value plots[65]. The largest box depicts the middle 50 % of the data. Each pair of subsequent smaller boxes depicts 50 % of the remaining data. The boxes are drawn until the estimated 95 % confidence intervals for the middle percentile in them would leak into the neighboring boxes. The remaining samples, indicated by dots, are consider outliers. The widths of the boxes are proportional to the number of samples inside them. The median is denoted by the black line. Traditional short- and long-range scaling exponents are shown in (a) and the optimal segmentation in (b). The scaling exponents are estimated from DFA1 fluctuation function.*

3–30 000 beat range. For such long scales detrending is increasingly important. Therefore explicit detrending is also considered, which is accomplished by determining the trend by a moving median filter with 101 beats wide kernel. The optimal segmentation with such detrending is illustrated in Fig. 4.11. The desirability plateaus at 3 segments but rises to the optimum at 5 segments. This may look confusing considering the pairs of nearly identical average slopes at the extremities of the fluctuation functions in Fig. 4.11(b). However, one must keep in mind that the segmentation simultaneously takes into account the individual variability of each fluctuation function, which may not be obvious in averages.

The distributions for the individual scaling exponents are presented in Fig. 4.12(a) for the actual RR intervals and in Fig. 4.12(b) for these detrended RR intervals. In (a) the shortest range (3–19 beats) scaling exponents are very similar to those of the traditional short-range alphas. However, the next range extends quite a bit further than the traditional long-range regime, and the separation between the groups is reduced. At even larger scales the groups remain approximately merged. The variance is also increased in the largest segment, possibly due to the statistical uncertainties in the fluctuation function. These findings support the hypothesis for the preeminence of the shorter scales. This is evident in the plummeting classification scores in Fig. 4.15(c), especially for CHF.

The situation is altered by the explicit detrending. The first two segments are very sim-

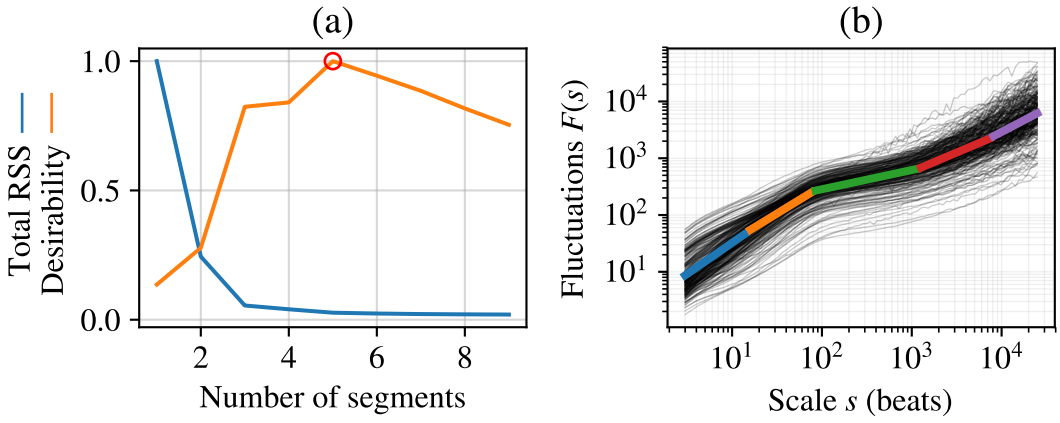

Figure 4.11. Optimal segmentation for detrended RR intervals in the 3–30 000 beat range. *The RR intervals are detrended by a moving median filter with 101 beats wide kernel. In (a) the total RSS and desirability are shown relative to their maximum values. In (b) the individual fluctuation functions are shown by the semi-transparent black lines. The thick colored lines depict the optimal segmentation, and their slopes are the averages of the corresponding slopes in the individual fluctuation functions. The fluctuations are computed by DFA1.*

ilar to the traditional short- and long-range results. In fact the classes are slightly more discernible in the corresponding larger scale (16-77 beats) segment. The detrending removes correlations in the mid-range (85–1156 beats) segment that includes the characteristic scale (101 beats) of the filtering kernel. At larger scales the correlations are restored but they remain stationary and the classes are more distinct than without detrending. This suggests that the detrending is successful at least to some extent. The classification results of Fig. 4.15(d) are quite similar to the two-range model, except that particularly the CHF cases become easier to distinguish.

The shared segmentation is useful for classification. However it could be beneficial to perform the analysis separately for different cardiac conditions to establish characteristic crossovers and segments. Additionally, individual segmentation could yield distributions for these traits.

The most interesting results are revealed by the full alpha spectra obtained by the Kalman smoother methods. The spectra for different orders of DFA are shown in Fig. 4.13. It is immediately clear that the non-stationary nature of the RR intervals cannot be analyzed by traditional non-detrending fluctuation analysis methods, including DFA0. Increasing the order of the DFA also increases the bias in small windows, and also the strength of the “oscillations” in the spectra are enhanced, along with variance. The peaks of the oscillations also shift slightly to larger scales as the order of DFA is increased. This distortion in DFA scale is theoretically predicted [30], which complicates the interpretation of crossovers in DFA. The spectra of CHF patients is particularly intriguing and two exponents are clearly not sufficient to fully describe it.

The number of patients is sufficient to attain accurate averages but individual variability

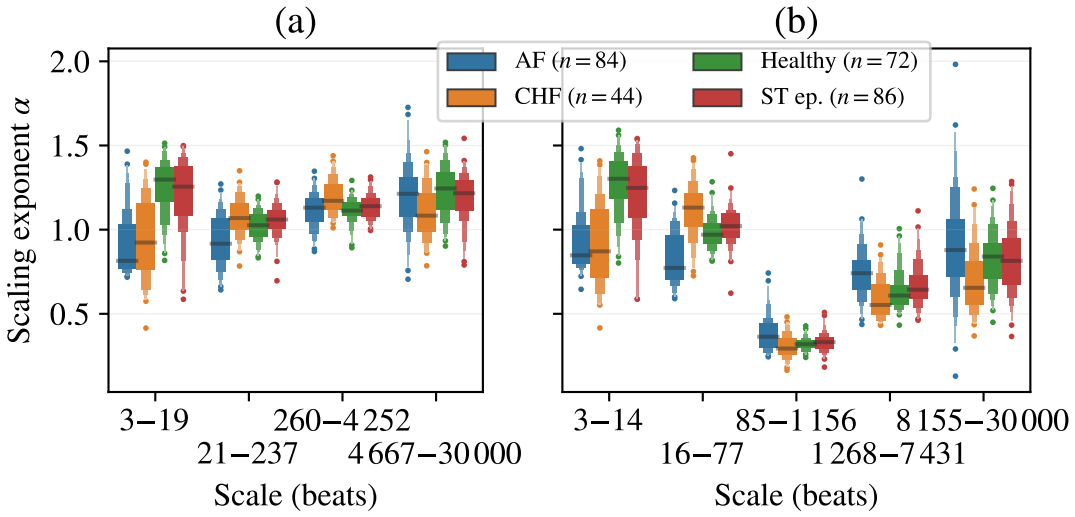

Figure 4.12. Scaling exponents of different cardiac conditions in the optimally segmented 3–30 000 beat range. *The distributions are illustrated by so-called letter-value plots[65]. The largest box depicts the middle 50 % of the data. Each pair of subsequent smaller boxes depicts 50 % of the remaining data. The boxes are drawn until the estimated 95 % confidence intervals for the middle percentile in them would leak into the neighboring boxes. The remaining samples, indicated by dots, are consider outliers. The widths of the boxes are proportional to the number of samples inside them. The median is denoted by the black line. In (a) the DFA1 fluctuation function is computed from the actual RR intervals and in (b) the RR intervals were explicitly detrended by a moving median filter with 101 beat wide kernel.*

is relatively large, which is expected to complicate classification and diagnosis. It should be noted that the average scaling is computed from the individual fluctuation functions of each patient. As was discussed in Sec. 2.2 this is a biased average. This may be justified by the observation in Fig. 4.11 that the individual fluctuation functions span almost two orders of magnitude at each scale. Therefore if a common fluctuation function was calculated the weighting for different patients would vary wildly.

Whether the analysis is performed by consider all beat-to-beat intervals or merely normal-to-normal interval is not found to significantly affect the average scaling behavior. Likewise, the mean alpha spectra are virtually equivalent for the 1st and 2nd order Kalman smoother models. However, *individual* spectra are smoother when utilizing the 2nd order model.

As it is uncertain whether the polynomial detrending in DFA is suitable for unknown trends, the RR intervals are also explicitly detrended by the moving median filter. The mean alpha spectra of the different cardiac conditions are illustrated in Fig. 4.14 with different kernel sizes for this filter. It is interesting to notice that the detrending removes long range correlations for scales roughly the size of the kernel. However, after this crossover point the different cardiac conditions become more clearly separated than without the detrending. This raises the question whether this is a real effect that has been hidden by

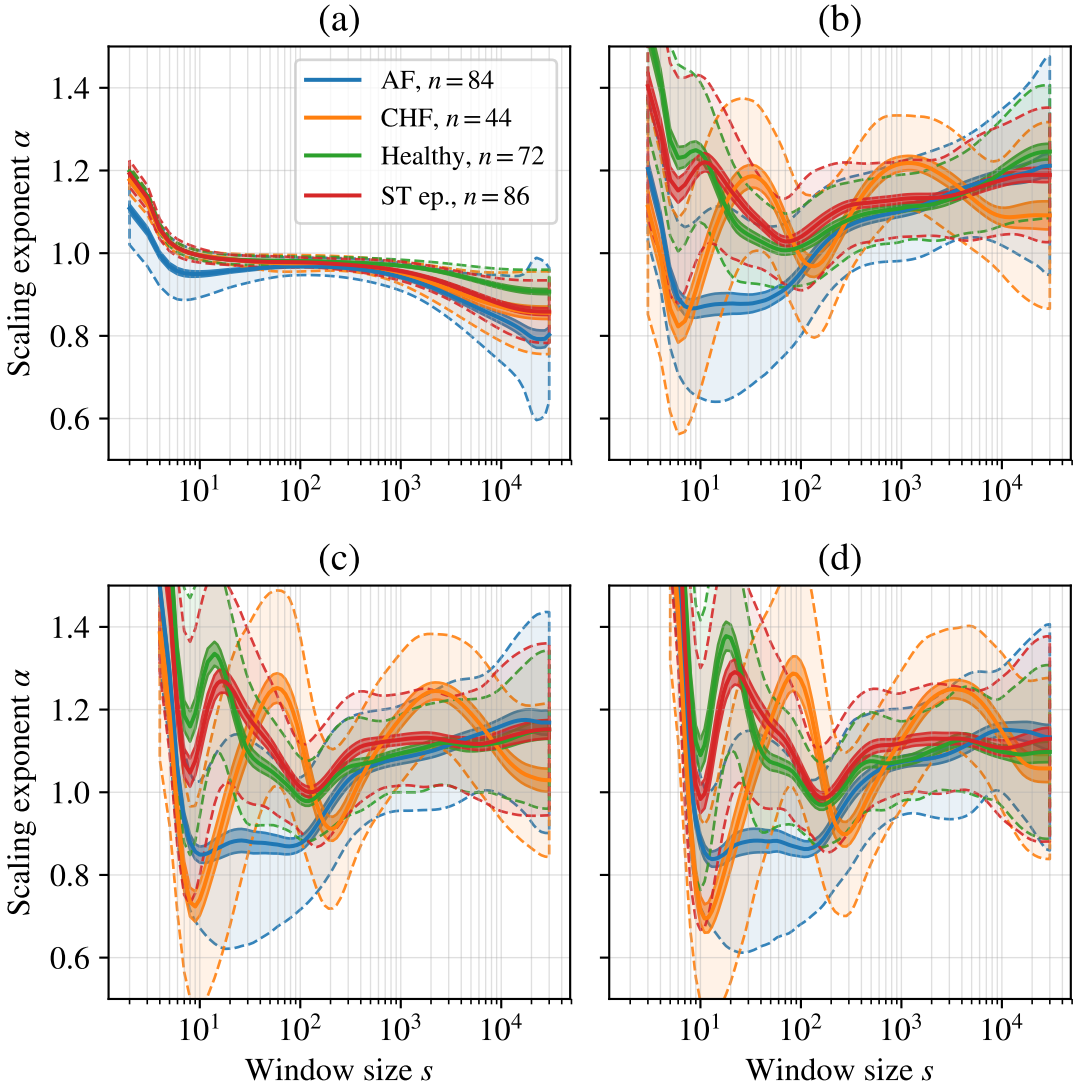

Figure 4.13. Alpha spectra of different cardiac conditions with different orders of DFA. *The darker shaded areas bounded by the solid lines denote the standard error of the mean, and the lighter regions enclosed by the dashed lines indicate the standard deviation. The order of the DFA is varied across the panels, and shown are results for DFA0 (a), DFA1 (b), DFA2 (c) and DFA3 (d).*

the non-stationarities in the RR intervals, or merely an artifact arising from the method.

The compelling consideration is whether these methods facilitate better classification of the cardiac conditions. Support vector machines with radial basis functions[64] provided overall the most consistent classification results from the considered algorithms. These results are summarized by Matthews correlation coefficients (MCC) in Fig. 4.15. For a binary classification problem the MCC is defined as

$$\mathrm{MCC} = \frac{T_0 T_1 - F_0 F_1}{\sqrt{(T_0 + F_0)\,(T_0 + F_1)\,(T_1 + F_0)\,(T_1 + F_1)}}, \tag{4.3}$$

where T_0 and T_1 are the numbers of correct predictions for the two classes, and similarly

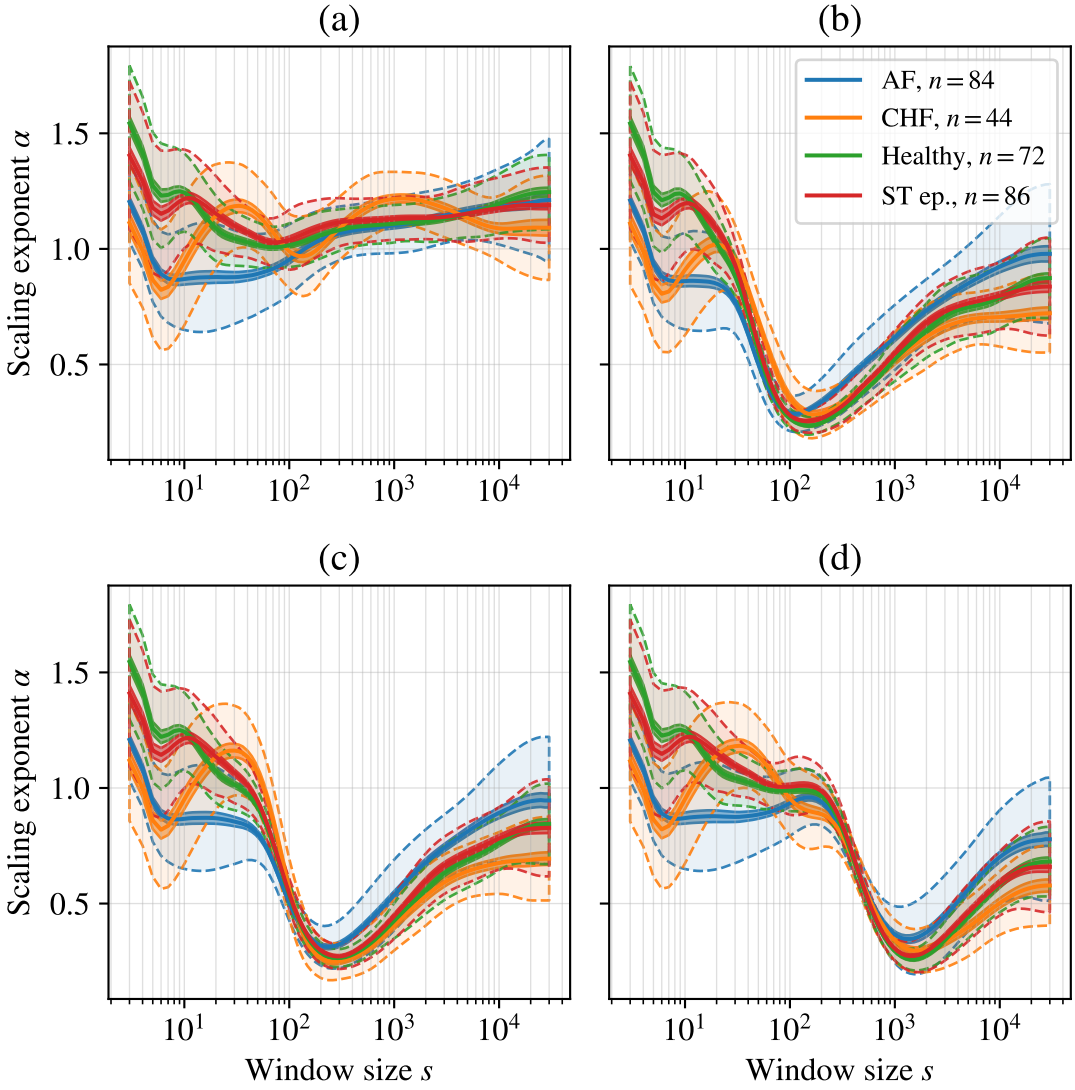

Figure 4.14. Alpha spectra of different cardiac conditions with explicitly detrended RR intervals. *The darker shaded areas bounded by the solid lines denote the standard error of the mean, and the lighter regions enclosed by the dashed lines indicate the standard deviation. In (a) the analysis has been performed on the actual RR intervals and in the other panels the RR intervals are first detrended by a moving median filter with a window lengths of 51 (b), 101 (c) and 501 (d) beats.*

F_0 and F_1 for their incorrect predictions. The MCC is limited to the interval between -1 and 1. A value of 1 means perfect classification, 0 implies no better than random prediction, and -1 indicates complete misclassification. The advantage of MCC over simpler statistics, such as the percentage of correctly classified samples (accuracy), is that it is a consistent measure even when the sizes of the two classes are imbalanced. [66, 67] While the MCC may be considered a good compromise, there is no perfect statistic for summarizing the full confusion matrix by a single number. As an example, the complete classification results by the 1st order Kalman smoother alpha spectra are shown in Fig. 4.16.

As can be seen in Fig. 4.15(a), the more complete analysis by the methods of this thesis suggests that even in the 3–64 beat range there is more information available than what is attainable by the traditional two-range scheme. Both of the new methods outperform the traditional method. On average the Kalman smoother alpha spectra surpasses the optimal segmentation, which is expected by the virtue of the spectra containing more features for classification. It varies across the classification pairs whether the 1st or 2nd order Kalman smoother model is better. This may be understood due to the smoother spectra of the 2nd order model occasionally suppressing noise and sometimes actual features.

When the explicit detrending is applied in Fig. 4.15(b) the results are somewhat mixed, as it benefits the classification for some pairs and hinders for others. Particularly the overall classification is enhanced for the optimal segmentation but deteriorated for the Kalman smoother methods. This may hint that for classification purposes the shorter range scaling is more important, as the density of features is larger at the longer scale for the Kalman smoother alpha spectra due to the logarithmic scale. This is further exacerbated by the mutual removal of long-range correlations close to the 100 beat regime due to the detrending.

Utilizing the whole range of 3–30 000 beats for classification reveals the intriguing observation in Fig. 4.15(c) that the results are actually worsened[2]. This corroborates the significance of the shorter scale for classification. However, employing the explicit detrending results in substantial improvement in Fig. 4.15(d). A captivating interpretation for this would be that the clearer distinction revealed by the detrending in the longest ranges of scaling in Fig. 4.14 is a real phenomenon. This is plausible as any extrinsic variations in the heart rate would result in non-stationarities which could be reasonably assumed to keep the scaling exponent slightly above one. Therefore the explicit detrending could be expected to reveal the intrinsic long-range correlations for the different cardiac conditions.

The relevance of the detrending, and the most appropriate method for doing so, remains an open research question. The results also strongly imply that certain ranges of scales are notably more important for classification, and consequently may be considered characteristic features of these cardiac conditions. While the shortest scale seems to be the most important, additional research is required to determine the most informative ranges of scaling. This kind of feature selection is a hot topic in machine learning [68].

As another example for future research, the new methods may also be utilized for the accurate study of the dynamical changes in the scaling behavior. An example of such an alpha landscape is shown in Fig. 4.17, which illustrates the daily fluctuations in the fractal scaling of heart rate variability. Physical exercise and sleep phases are discernible.

[2]The short- and long-range scaling exponents are still extracted from the same ranges as before, so those results remain the same, except for small variations due to less window sizes being present for the linear regression instead of the full 4–64 beat range.

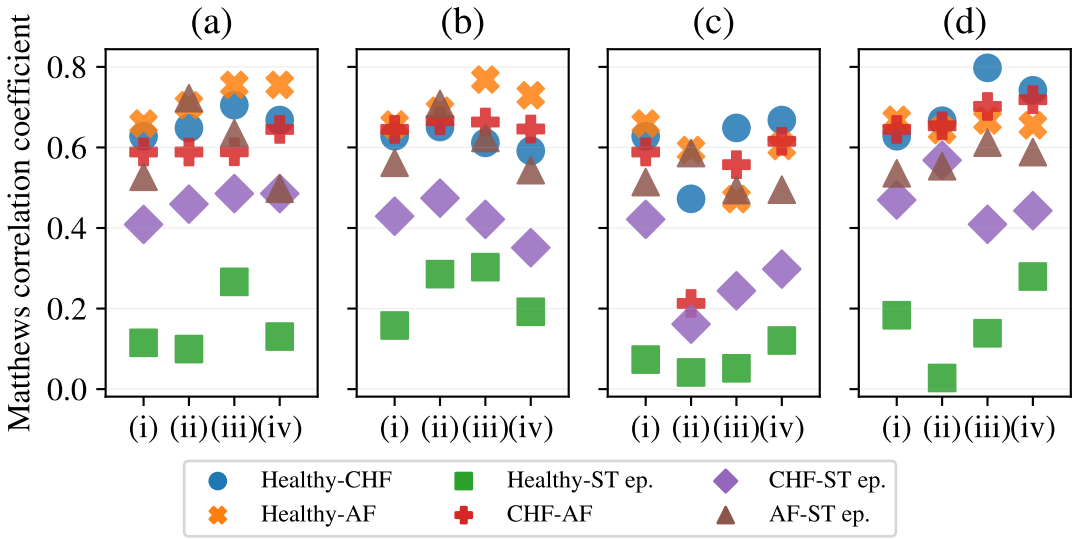

Figure 4.15. Matthews correlation coefficients for binary classification of cardiac conditions by their fractal scaling. *The classification is performed by support vector machines with radial basis functions and the hyperparameters are optimized by a simple grid search. Different methods for estimating the scaling exponents are shown on the horizontal axis as follows: traditional short- and long-range (i), optimal linear segmentation (ii), 1st order Kalman smoother (iii), and 2nd order Kalman smoother (iv). The panels (a) and (b) show the results when the maximum scale is limited to 64 beats for consistency with the traditional short- and long-range scaling exponents, and in (b) the explicit detrending is performed prior to the DFA by the moving median filter with 101 beats wide kernel. In panels (c) and (d) the maximum scale is 30 000 beats, and the explicit detrending is again employed in (d).*

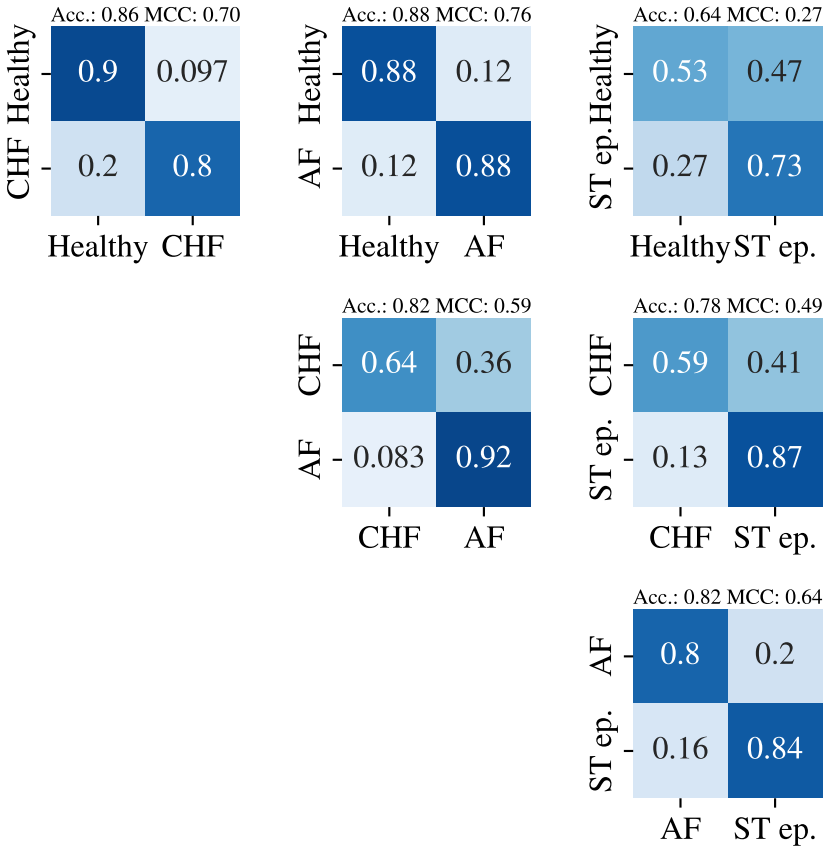

Figure 4.16. Confusion matrices for binary classification of cardiac conditions by their fractal scaling. *The features utilized for the classification are the alpha spectra obtained by the 1st order Kalman smoother method in the 3–64 beat range. The results are illustrated as relative confusion matrices where the rows correspond to the correct classes and the columns are the predicted classes. Above each matrix also the accuracy (Acc.) and the Matthews correlation coefficient (MCC) are shown.*

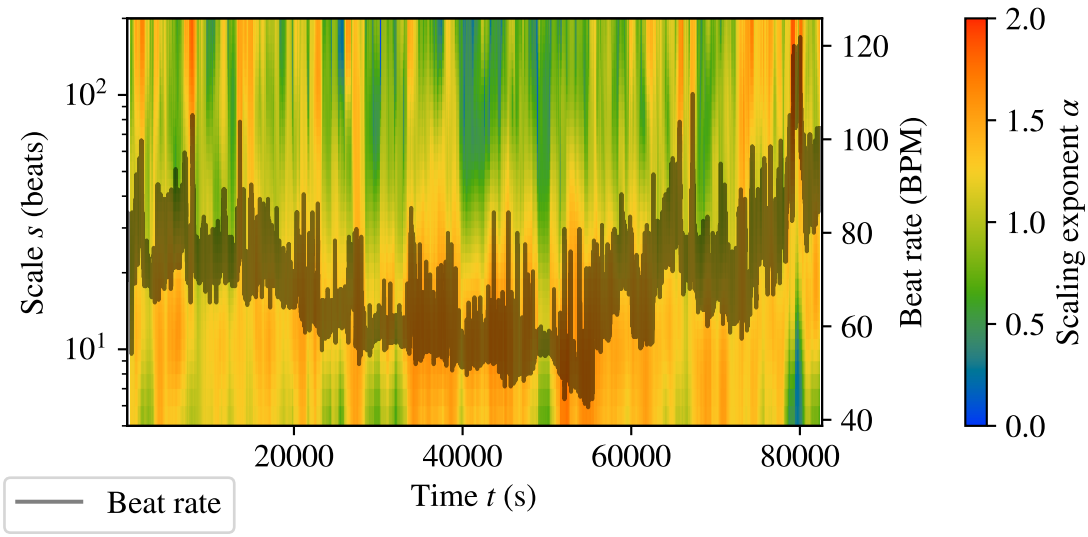

Figure 4.17. Example of an alpha landscape. *The daily fluctuations in the fractal scaling of RR intervals are shown by performing DFA1 in 1000 beat windows and extracting the alpha spectra by the 1st order Kalman smoother method.*

5 CONCLUSIONS

The methods developed and described in this thesis are found to be useful tools for deciphering the DFA scaling exponents. Particularly the alpha spectra obtained by the Kalman smoother is useful for uncovering details in the scaling behavior that would remain hidden from simple linear regression estimates. The advantages of this Kalman smoother alpha spectra estimator over earlier approaches, such as local-scale exponents [14] and the alpha-beta filter [13], are its parameter-free implementation and robustness in the presence of statistical noise in the fluctuation functions.

The method for finding optimal linear segmentations is practical when complete alpha spectra would be largely redundant and only a few point estimates would suffice. Other segmentation methods exist that provide mathematically more rigorous approach to determining crossovers in the fluctuation function [16]. The versatility of the linear integer programming framework may, however, be beneficial by providing easily customizable approach that can be adapted to include additional constraints. An example of this is already presented, as the method is readily generalized to handle groups of fluctuation functions to provide globally optimal segmentations. This is attractive for, e.g., acquiring mutually consistent feature vectors for machine learning applications. The optimization procedure could also be modified by attempting to maximize the Matthews correlation coefficients of the classification task instead of minimizing for the squared residuals of linear regression. This could be more beneficial from diagnosis and feature selection point of view.

It is also clear that the fractal scaling of RR intervals is not sufficient for distinguishing all cardiac conditions, such as ST episodes. Therefore other time series from ECGs could also be exploited for the analysis, such as QT or ST intervals. Such approaches could also benefit from additional investigation by detrended cross-correlation analysis [69] or transfer entropy studies [70]. The classification of the heart diseases in this thesis is largely exploratory to demonstrate the viability of the new methods. A far more sophisticated foray into this field is warranted. Integrating the methods of this thesis into an all-encompassing cardiac health diagnostics package is an ambition for tomorrow.

After establishing its validity, the Kalman smoother method for the alpha spectra may be utilized for various purposes in future studies. For instance, it may be employed for extracting temporal landscapes of the scaling exponent both as a function of the scale and time. As was seen in Fig. 4.17, physical exercise leaves a distinct mark in the scaling. This is encouraging for applications in sports analytics. Another possibility arises in the

feasibility of obtaining better estimates for the Hurst surfaces of multiscale multifractal analysis [41].

Previous studies may also have overlooked insights hidden in the details that would be revealed by the techniques of this thesis. It was noticed that patients suffering from congestive heart failure exhibit peculiar scaling in their heart rate variability that warrants further research. The results also suggest that detrending in the context of HRV could benefit from additional study. Specifically, it would appear that explicit detrending of the RR intervals enhances the discernibility of long range correlations in different cardiac conditions. This alludes that the piecewise polynomial detrending embedded in DFA is not sufficient for eliminating external trends in the heart rate.

The alpha spectra is an invaluable tool for verification of simulated results when combined with the analytical expressions [21] for the expected DFA fluctuation function. This provided key insight into the characteristics of different time series generation methods. Proper care must be taken if conclusions are to be drawn from numerical experiments with simulated data. It is plausible that the limitations of Fourier filtering have contributed to the belief that DFA is not particularly accurate for highly anticorrelated time series.

The theoretical results of Ref. [21] may be useful starting point for obtaining analytical expressions for the variance of the DFA fluctuation function, and expanding them for the detrended multifractal and cross-correlation formalisms could be fruitful. Current analysis could already be enhanced by studying the expected behavior of the fluctuation function across all the scales for different processes. Combined with the alpha spectra, it may be possible to devise an iterative correction for the bias in the fluctuation function. Another possibility is to change the viewpoint and consider the inverse problem: Employing robust estimation procedure for the autocovariance function from the fluctuation function. However, this is not trivial, as it is ill-posed problem that requires regularization due to the corresponding linear system being underdetermined.

It is worth noting that while this thesis focuses on applications in computational cardiology, the methods themselves are generally applicable to time series analysis, and even more universally to any tasks that depend on estimating local slopes of some variable. In particular DFA is intimately linked to diffusion, and the enhanced analysis aided by the theoretical understanding of the fluctuation function may provide a robust framework for the investigation of the omnipresent diffusion processes.

Perhaps the most important conclusion to be drawn from experimenting with the alpha spectra is that the details of the fractal scaling as a function of the scale are important, and that these details may be neglected if only considering asymptotic behavior. The analysis so far has barely scratched the surface and raises more interesting research questions than it has been able to answer. As a final note, the implemented methods will be published as open source software at a later date in accordance to the principles of open science.

REFERENCES

[1] B. B. Mandelbrot, *The fractal geometry of nature*, English, Rev. (Freeman, New York, 1982).

[2] C. 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 **5**, 82–87 (1995).

[3] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng and H. E. Stanley, Fractal dynamics in physiology: alterations with disease and aging, Proceedings of the National Academy of Sciences **99**, 2466–2472 (2002).

[4] H. E. Hurst, The problem of long-term storage in reservoirs, International Association of Scientific Hydrology. Bulletin **1**, 13–27 (1956).

[5] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley and A. L. Goldberger, Mosaic organization of dna nucleotides, Physical Review E **49**, 1685–1689 (1994).

[6] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. Rego, S. Havlin and A. Bunde, Detecting long-range correlations with detrended fluctuation analysis, Physica A: Statistical Mechanics and its Applications **295**, 441–454 (2001).

[7] M. S. Santhanam, J. N. Bandyopadhyay and D. Angom, Quantum spectrum as a time series: fluctuation measures, Phys. Rev. E **73**, 015201 (2006).

[8] N. Vandewalle and M. Ausloos, Coherent and random sequences in financial fluctuations, Physica A: Statistical Mechanics and its Applications **246**, 454–459 (1997).

[9] E. Räsänen, O. Pulkkinen, T. Virtanen, M. Zollner and H. Hennig, Fluctuations of hi-hat timing and dynamics in a virtuoso drum track of a popular music recording, PLOS ONE **10**, 1–16 (2015).

[10] R. E. Kalman, A new approach to linear filtering and prediction problems, English, Journal of Basic Engineering **82**, 35–45 (1960).

[11] H. E. Rauch, F. Tung and C. T. Striebel, Maximum likelihood estimates of linear dynamic systems, AIAA Journal **3**, 1445–1450 (1965).

[12] G. M. Viswanathan, C.-K. Peng, H. E. Stanley and A. L. Goldberger, Deviations from uniform power law scaling in nonstationary time series, Phys. Rev. E **55**, 845–849 (1997).

[13] J. C. Echeverría, M. S. Woolfson, J. A. Crowe, B. R. Hayes-Gill, G. D. H. Croaker and H. Vyas, Interpretation of heart rate variability via detrended fluctuation analysis and $\alpha\beta$ filter, English, Chaos: An Interdisciplinary Journal of Nonlinear Science **13**, 467–475 (2003).

[14] P. Castiglioni, G. Parati, A. Civijian, L. Quintin and M. D. Rienzo, Local scale exponents of blood pressure and heart rate variability by detrended fluctuation analysis:

effects of posture, exercise, and aging, IEEE Transactions on Biomedical Engineering **56**, 675–684 (2009).

[15] J. Xia, P. Shang and J. Wang, Estimation of local scale exponents for heartbeat time series based on DFA, Nonlinear Dynamics **74**, 1183–1190 (2013).

[16] E. Ge and Y. Leung, Detection of crossover time scales in multifractal detrended fluctuation analysis, Journal of Geographical Systems **15**, 115–147 (2013).

[17] A. Habib, J. P. Sorensen, J. P. Bloomfield, K. Muchan, A. J. Newell and A. P. Butler, Temporal scaling phenomena in groundwater-floodplain systems using robust detrended fluctuation analysis, Journal of Hydrology **549**, 715–730 (2017).

[18] M. Molkkari and E. Räsänen, Robust estimation of the scaling exponent in detrended fluctuation analysis of beat rate variability, Computing in cardiology (2018).

[19] Z. Chen, P. C. Ivanov, K. Hu and H. E. Stanley, Effect of nonstationarities on detrended fluctuation analysis, Phys. Rev. E **65**, 041107 (2002).

[20] M. Höll and H. Kantz, The relationship between the detrendend fluctuation analysis and the autocorrelation function of a signal, The European Physical Journal B **88**, 327 (2015).

[21] O. Løvsletten, Consistency of detrended fluctuation analysis, Phys. Rev. E **96**, 012141 (2017).

[22] G. E. Billman, H. V. Huikuri, J. Sacha and K. Trimmel, An introduction to heart rate variability: methodological considerations and clinical applications, Frontiers in Physiology **6**, 55 (2015).

[23] C. Chatfield, *The analysis of time series: an introduction*, English, 5th (Chapman and Hall/CRC, 1996).

[24] B. Mandelbrot and J. Van Ness, Fractional brownian motions, fractional noises and applications, SIAM Review **10**, 422–437 (1968).

[25] C. Heneghan and G. McDarby, Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes, Phys. Rev. E **62**, 6103–6110 (2000).

[26] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons and H. E. Stanley, Long-range correlations in nucleotide sequences, Nature **356**, 168–170 (1992).

[27] J. W. Kantelhardt, Fractal and multifractal time series, *Mathematics of complexity and dynamical systems*, edited by R. A. Meyers (Springer New York, New York, NY, 2011), 463–487.

[28] M. Höll, H. Kantz and Y. Zhou, Detrended fluctuation analysis and the difference between external drifts and intrinsic diffusionlike nonstationarity, Phys. Rev. E **94**, 042201 (2016).

[29] R. M. Bryce and K. B. Sprague, Revisiting detrended fluctuation analysis, Scientific reports **2**, 315 (2012).

[30] K. Kiyono, Establishing a direct connection between detrended fluctuation analysis and fourier analysis, Phys. Rev. E **92**, 042925 (2015).

[31] Q. Yuan, C. Gu, T. Weng and H. Yang, Unbiased detrended fluctuation analysis: long-range correlations in very short time series, Physica A: Statistical Mechanics and its Applications **505**, 179–189 (2018).

[32] A. Carbone, Detrending moving average algorithm: a brief review, 2009 ieee toronto international conference science and technology for humanity (tic-sth) (Sept. 2009), 691–696.

[33] X.-Y. Qian, G.-F. Gu and W.-X. Zhou, Modified detrended fluctuation analysis based on empirical mode decomposition for the characterization of anti-persistent processes, English, Physica A: Statistical Mechanics and its Applications **390**, 4388–4395 (2011).

[34] E. Maiorino, F. M. Bianchi, L. Livi, A. Rizzi and A. Sadeghian, Data-driven detrending of nonstationary fractal time series with echo state networks, Information Sciences **382-383**, 359–373 (2017).

[35] A. Clauset, C. Shalizi and M. Newman, Power-law distributions in empirical data, SIAM Review **51**, 661–703 (2009).

[36] J. Echeverria, E. Rodriguez, M. Aguilar-Cornejo and J. Alvarez-Ramirez, Linear combination of power-law functions for detecting multiscaling using detrended fluctuation analysis, Physica A: Statistical Mechanics and its Applications **460**, 283–293 (2016).

[37] A. Khintchine, Korrelationstheorie der stationären stochastischen prozesse, Mathematische Annalen **109**, 604–615 (1934).

[38] D. ben-Avraham and S. Havlin, *Diffusion and reactions in fractals and disordered systems*, English (Cambridge University Press, Jan. 2005).

[39] J. Feder, *Fractals (physics of solids and liquids)*, English (Springer, 1988).

[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, English, Physica A: Statistical Mechanics and its Applications **316**, 87–114 (2002).

[41] J. Gierałtowski, J. J. Żebrowski and R. Baranowski, Multiscale multifractal analysis of heart rate variability recordings with a large number of occurrences of arrhythmia, Phys. Rev. E **85**, 021915 (2012).

[42] S. Särkkä, *Bayesian filtering and smoothing*, English (Cambridge University Press, Cambridge, 2013).

[43] B. Fornberg, Generation of finite difference formulas on arbitrarily spaced grids, Mathematics of Computation **51**, 699–706 (1988).

[44] A. Schrijver, *Theory of linear and integer programming*, English (Wiley, Chichester, 1986).

[45] R. Goldsmith, Congestive heart failure RR interval database, physionet.org (2003) 10.13026/c2f598 (visited on 01/28/2019).

[46] D. Saupe, The science of fractal images, edited by H.-O. Peitgen and D. Saupe (Springer, New York, 1988) Chap. Algorithms for random fractals.

[47] H. A. Makse, S. Havlin, M. Schwartz and H. E. Stanley, Method for generating long-range correlations for large systems, Phys. Rev. E **53**, 5445–5449 (1996).

[48] R. B. Davies and D. S. Harte, Tests for hurst effect, Biometrika **74**, 95–101 (1987).

[49] A. Dieker and M. Mandjes, On spectral simulation of fractional brownian motion, Probability in the Engineering and Informational Sciences **17**, 417–434 (2003).

[50] Y.-H. Shao, G.-F. Gu, Z.-Q. Jiang, W.-X. Zhou and D. Sornette, Comparing the performance of fa, dfa and dma using different synthetic long-range correlated time series, Scientific Reports **2**, 835 (2012).

[51] M. Malik et al., Heart rate variability: standards of measurement, physiological interpretation, and clinical use, European Heart Journal **17**, 354–381 (1996).

[52] J.-G. Dong, The role of heart rate variability in sports physiology (review), Experimental and Therapeutic Medicine **11**, 1531–1536 (2016).

[53] A. L. Goldberger, Z. D. Goldberger and A. Shvilkin, *Goldberger's clinical electrocardiography: a simplified approach*, English, 9th (Elsevier, 2017).

[54] P. Castiglioni et al., Assessing the fractal structure of heart rate by the temporal spectrum of scale exponents: a new approach for detrended fluctuation analysis of heart rate variability, Biomedizinische Technik. Biomedical engineering **56**, 175–83 (2011).

[55] The Arrhythmia Laboratory At Boston's Beth Israel Hospital (Now The Beth Israel Deaconess Medical Center), The MIT-BIH normal sinus rhythm database, physionet.org (1990) 10.13026/c2nk5r (visited on 01/28/2019).

[56] A. L. Goldberger et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation **101**, e215–e220 (2000).

[57] P. Stein and R. Goldsmith, Normal sinus rhythm RR interval databse, physionet.org (2003) 10.13026/c2s881 (visited on 01/28/2019).

[58] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman and E. Braunwald, The BIDMC congestive heart failure database, physionet.org (2000) 10.13026/c29g60 (visited on 01/28/2019).

[59] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman and E. Braunwald, Survival of patients with severe congestive heart failure treated with oral milrinone, Journal of the American College of Cardiology **7**, 661–670 (1986).

[60] S. Petrutiu, A. V. Sahakian and S. Swiryn, The long-term AF database, physionet.org (2003) 10.13026/c2qg6q (visited on 01/28/2019).

[61] A. V. Sahakian, S. Petrutiu and S. Swiryn, Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans, EP Europace **9**, 466–470 (2007).

[62] F. Jager, A. Taddei, G. B. Moody, M. Emdin, G. Antolič, R. Dorn, A. Smrdel, C. Marchesi and R. G. Mark, The long-term ST database, physionet.org (2007) 10.13026/C2G01T (visited on 01/28/2019).

[63] F. Jager, A. Taddei, G. B. Moody, M. Emdin, G. Antolič, R. Dorn, A. Smrdel, C. Marchesi and R. G. Mark, Long-term ST database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia, Medical and Biological Engineering and Computing **41**, 172–182 (2003).

[64] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, Scikit-learn: machine learning in Python, Journal of Machine Learning Research **12**, 2825–2830 (2011).

[65] H. Hofmann, H. Wickham and K. Kafadar, Letter-value plots: boxplots for large data, Journal of Computational and Graphical Statistics **26**, 469–477 (2017).

[66] B. Matthews, Comparison of the predicted and observed secondary structure of t4 phage lysozyme, Biochimica et Biophysica Acta (BBA) - Protein Structure **405**, 442–451 (1975).

[67] S. Boughorbel, F. Jarray and M. El-Anbari, Optimal classifier for imbalanced data using Matthews correlation coefficient metric, PLOS ONE **12**, 1–17 (2017).

[68] I. Guyon and A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research **3**, 1157–1182 (2003).

[69] B. Podobnik and H. E. Stanley, Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series, Phys. Rev. Lett. **100**, 084102 (2008).

[70] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. **85**, 461–464 (2000).