Anomalous statistical properties and fluctuations on multiple timescales

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

vorgelegt

dem Bereich Mathematik und Naturwissenschaften

der Technischen Universität Dresden

von

M.Sc., Philipp Gerhard Meyer

geboren am 06.12.1991 in Fürth

Eingereicht am 17.12.2019

Gutachter:

Prof. Dr. Holger Kantz
Max-Planck-Institute for the Physics of Complex Systems
Noethnitzer Str. 38
01187 Dresden
Germany

Prof. Dr. Kevin E. Bassler
University of Houston
617 Science and Research Bld. 1
Houston, Texas 77204-5005
USA

Prof. Dr. Marc Timme
Technical University of Dresden
Helmholtzstr. 18
01069 Dresden
Germany

Danksagung

Ich möchte allen Danken, die mich in den letzten Jahren bei meiner Promotion unterstützt haben. Vor allem Holger Kantz, der mir diese Möglichkeit gegeben hat. Vielen Dank, dass ich an diesen interessanten Themen arbeiten durfte und für die große Unterstützung, die vielen Vorschläge und Ideen. Auch allen anderen, die mir in Diskussionen neue Sichtweisen und Methoden nähergebracht haben, möchte ich danken. Besonders hervorheben möchte ich dabei diejenigen, mit denen ich an verschiedenen Projekten zusammengearbeitet habe. Eli Barkai, Marc Höll, Kevin Bassler, Mehrnaz Anvari und Erez Aghion haben mich in an neue Ideen und Fragen herangeführt. All diese Projekte haben viel Spaß gemacht. Ich habe die Atmosphäre in der Gruppe Nonlinear time series analysis immer sehr genossen. Dafür danke ich allen, die in den letzten Jahren Mitglieder der Gruppe waren. Ich danke auch den Gutachtern und sonstigen Mitgliedern des Promotionskomitees dafür, dass sie sich die Zeit nehmen sich mit meiner Arbeit auseinander zu setzen. Auch im privaten kann ich nur allen für die letzten drei Jahre danken, sowohl Freunden, als auch meiner Familie. Ganz besonders danke ich meinen Eltern, die mich schon immer unterstützt haben.

Zusammenfassung

Wie kann man Fluktuationen in eindimensionalen Zeitreihen charakterisieren und wie können gemessene Effekte in ihre dynamischen Ursachen zerlegt werden? Im Kontext dieser Fragen werden eine Reihe von unterschiedlichen Problemen diskutiert und Lösungen vorgestellt.

Zunächst werden die Ursachen anomaler Diffusion behandelt. Ein kürzlich vorgeschlagener Ansatz ist die Zerlegung des Hurst Exponenten in den Joseph-, Noah-, und Moseseffekt. Diese repräsentieren die Verletzungen der drei Bedingungen des zentralen Grenzwertsatzes. In dieser Arbeit wird das Konzept auf ein intermittentes deterministisches System angewandt, welches eine Kombination aller drei Effekte zeigt. Dennoch kann man anhand des Ergebnisses die Dynamik intuitiv verstehen. Zusätzlich wird das Konzept theoretisch untersucht und mit einer Herleitung in Verbindung gebracht, welche Gültigkei des Ansatzes für eine allgemeine Klasse von Systemen zeigt.

Wenn erst einmal der Typ der anomalen Statistik klassifiziert ist, kann man als nächstes nach der dynamischen Ursache des Effekts fragen. Insbesondere die Eigenschaft langreichweitiger Korrelationen in der Zeit (dem Josepheffekt) wird hier genauer diskutiert. In Messungen können sie aufgrund unterschiedlicher dynamische Ursachen oder als emergentes Phänomen auftreten. Hier wird eine Sammlung verschiedener Erklärungen vorgestellt.

Eine beliebte Methode zur Erkennung langreichweitiger Korrelationen ist Detrended Fluctuation Analysis (DFA). Ihre Vorteile gegenüber traditioneller Methoden sind Stabilität und Gleichmäßigkeit bis zu Zeitskalen um ein Viertel der Messzeit, sowie die Vernachlässigung von langsamer Dynamik und Trends. Kürzlich wurde eine analytische Theorie der Methode vorgestellt. In dieser Dissertation wird die Methode weiterentwickelt und analysiert. Es wird ein Ansatz vorgestellt, wie DFA auch kurzreichweitig korrelierte Daten und sogar sehr komplexe Dynamik beschreiben kann. Die Eigenschaften können durch eine Zerlegung in eine Superposition linearer Modelle erklärt werden.

Diese Dissertation dient dem Verständnis anomaler statistischer Eigenschaften. Zusätzlich wird eine dafür verwendete Methode weiterentwickelt und auf normale und komplexe Dynamik angewendet. Die Atmosphäre ist ein gutes Beispiel für solch ein komplexes System, bei dem die vorgestellten stochastischen Methoden nützlich sind. Hier wird gezeigt, dass mithilfe der DFA Schwingungsmoden erkannt und Periodendauern bestimmt werden können. Eine dieser Moden ist die El Niño Southern Oscillation. Eine weniger Bekannte und auch schwierigere Anwendung bietet eine Periode von 7–8 Jahren bei Temperaturschwankungen in Europa. Eine ganz andere Art von komplexen Systemen sind Stromnetze. Dennoch kann man mit der neuen Methode auch hier Datenmodelle generieren, die die wichtigsten Eigenschaften von Stromnetzfrequenzen reproduzieren.

Abstract

How can fluctuations in one-dimensional time series data be characterized and how can detected effects be decomposed into their dynamical origins or causes? In the context of these questions, a variety of problems are discussed and solutions are introduced.

The first issue concerns the causes of anomalous diffusion. A previously proposed framework decomposes the Hurst exponent into the Joseph, Noah, and Moses effects. They represent violations of the three premises of the central limit theorem. Here the framework is applied to an intermittent deterministic system, which exhibits a rich combination of all three effects. Nevertheless, the results provide an intuitive interpretation of the dynamics. In addition, the framework is theoretically discussed and connected to a calculation that proves its validity for a large class of systems.

Once the type of anomalous statistical behavior is classified, one might ask what the dynamical origin of the effects is. Especially the property of long range temporal correlations (the Joseph effect) is discussed in detail. In measurements, they might arise from different dynamical origins or can be explained as an emerging phenomenon. A collection of different routes to the observed behavior is established here.

A popular tool for detecting long range correlations is detrended fluctuation analysis. Its advantages over traditional methods are stability and smoothness for timescales up to one fourth of the measurement time and the ability to neglect the slow dynamics and trends. Recently, a theory for an analytical understanding of this method was introduced. In this thesis, the method is further analyzed and developed. An approach is presented that enables scientists to use this method for short range correlated data, even if the dynamics is very complex. Fluctuations can be decomposed into a superposition of linear models that explain its features.

Therefore, on the one hand, this thesis is about understanding the effects of anomalous diffusion. On the other hand, it is about widening the applicability of one of its detection methods such that it becomes useful for understanding normal or complex statistical behavior. A good example of a complex system, where the proposed stochastic methods are useful, is the atmosphere. Here it is shown how detrended fluctuation analysis can be used to uncover oscillatory modes and determine their periods. One of them is the El Niño southern oscillation. A less well known and more challenging application is a 7–8 year mode in European temperature fluctuations. A power grid is a very different type of complex system. However, using the new method, it is possible to generate a data model that incorporates the important features of the grid frequency.

Contents

1 Introduction

This work is about data that, like all measured data, fluctuates in time. The reason might be noise due to external perturbations, the chaotic nature of the underlying dynamics or simply measurement noise. The characterization of these fluctuations enables scientists to model the system with a stochastic process, to compare different recordings and to determine uncertainties of predictions. Identifying regular patterns enables predictions of the future of the system. A data model is a model that reproduces statistical properties of a complex system despite being uninformed about the underlying physical processes. It does not require deep physical understanding, which is often not available due to the complexity of the system.

The emphasis in this thesis is on anomalous statistical properties that violate the premises of the central limit theorem. In temperature time series, for example, such a property was detected by the method of detrended fluctuation analysis [39]. Several questions arise from such a finding. Since there is more than one way to violate the premises of the central limit theorem, the first question is: (1) which property does the method really detect? In order to answer this question, a framework is needed, which intuitively decomposes anomalous statistical effects. The answer in the case of detrended fluctuation analysis is that the method detects long range correlations [65]. This means intuitively that the present state of the system depends on its complete history [49]. However, this answer leads to the next question: (2) what are the causes of long range correlations, i.e., how is it generated dynamically or can it also pop up as a statistical artifact? A third question concerning the detected long range correlations in temperature time series is: (3) what is the connection between this finding and the oscillatory modes and patterns that were also detected in the same datasets [99]?

The answers (or potential answers) to all three questions are discussed in this thesis. The focus is on one-dimensional time series both from models and measured data. The models can be stochastic or chaotic. In this context, this thesis covers different topics starting from a very general framework of anomalous statistical properties. From these general considerations, more specific topics are discussed. One of the anomalous properties and one popular detection method (detrended fluctuation analysis) are investigated in more detail. The applications are then mostly taken from one specific field, which is atmospheric science. The following sections introduce these different layers, starting with the abstract topic of anomalous statistical properties going towards the data analysis problem of climate variability.

1.1 Anomalous statistical properties

Anomalous statistical behavior here refers to statistics that does not follow the central limit theorem and the law of large numbers. It has been studied for a long time and was detected in many time series. For some systems, people have already claimed that 'anomalous is normal' [109]. Popular detection methods like detrended fluctuation analysis [97] were cited in thousands of articles. Accordingly, there is a large amount of literature that shows anomalous statistical behavior in real-world data from all kinds of sources. Market volatility [77] and turbulent flows [103] are just two old and popular examples. Anomalous effects were observed, for instance, in blinking quantum dots [22], laser cooling [14], and human heartbeats (in some phases) [24].

Anomalous diffusion is a consequence of anomalous statistical properties of the increments or velocity of the respective system. The dynamical origins are described by a large number of popular models for one-dimensional systems. Some of them are still subject of very active research despite not looking too complex at first sight. However, the anomalous nature of their dynamics makes analytical understanding hard in some cases. Examples for such systems are fractional Brownian motion [78], Lévy walks [117], scaled Brownian motion [63], and continuous time random walks [90]. Some of these models are more than 50 years old. However, they are subjects of research until today [86].

Anomalous diffusion is defined by the scaling of the mean squared displacement and is usually also measured in this way [86]. The ensemble average of the mean squared displacement scales in time with the exponent of $2H$, where, in contrast to normal diffusion [37], the Hurst exponent is $H \neq 1/2$. However, sometimes the Hurst exponent is also measured with other methods, like detrended fluctuation analysis [97] or recently machine learning algorithms [19]. The problem is that these methods are not always equivalent, which leads to confusion if results are compared. An example is the history of investigations on intra-day market returns in finance [16, 114, 27]. Here the Hurst exponent is due to non-stationarity of the increments and can therefore not be detected using detrended fluctuation analysis, which is a measure for correlations.

Even if people could agree on the mean squared displacement as the definition of the Hurst exponent, this would still not be a good solution for time series analysis of data. The class of systems that exhibit anomalous Hurst exponents includes many sub-classes that do not really seem to be related at second sight. While long range correlations can cause anomalous diffusion with some exponent $H \neq 1/2$, the same exponent could also be generated by a series of completely uncorrelated values, if their probability distribution had very fat tails.

It is, therefore, useful to have an understandable framework in place that is able to classify the effects that lead to anomalous diffusion in a meaningful way. For practitioners, this would be beneficial when comparing the result of some data analysis to the numerous models of anomalous diffusion that have been suggested. Of course, the framework then has to be understood in depth in order to not lead to additional confusion.

The idea of decomposing anomalous statistical effects was already introduced by Man-

delbrot and Wallis [79] very early in the history of anomalous diffusion. In the context of hydrology, they wanted to describe that 'a long period of unusual (high or low) precipitation can be extremely long' but also 'extreme precipitation can be very extreme indeed'. Long range correlations (i.e. diverging correlation times) lead to the so-called Joseph effect, while fat tails (i.e. diverging second moment) of the probability distribution lead to a Noah effect. The framework was extended by including non-stationary increments as the Moses effect by Chen and coworkers [27], keeping with the biblically themed names. The three effects are represented by exponents that are linked to the Hurst exponent by a scaling relation.

While the decomposition of the Hurst exponent was already proposed, the contribution of this work is mostly to increase the understanding of the scaling relations and the effects. Firstly, the postulated scaling relation is linked to an analytical derivation that shows its validity for a large class of systems. Secondly, the behavior of the exponents is examined in systems that fundamentally differ from the paradigmatic models considered in [27]. The effects of infinite invariant densities are discussed as well as other cases that push the theory to its limits. It turns out that the proposed interpretation of the exponents remains valid in these cases. Thirdly, the decomposition is performed for a deterministic dynamical system, which exhibits a rich mixture of all effects. Here the framework helps to intuitively understand the dynamics.

1.2 Long range correlations

One of the anomalous statistical effects introduced above is of special interest in this thesis. This is the Joseph effect that describes long range temporal correlations, also called 'long memory' [49]. Long range correlations describe dynamics with autocorrelations that typically decay with power laws. This leads to situations where no characteristic time can be defined for which correlations decay. Intuitively this means that 'there is non-negligible dependence between the present and all points in the past' [49]. It is the same property that Hurst detected in time series of river Nile basin levels [60]. It is traditionally measured by rescaled range (R/S) statistics.

A more modern method is detrended fluctuation analysis. It is discussed in this thesis that for most (but not for all) processes the result of this method is equivalent to the Joseph exponent. The method generates the so-called fluctuation function, which is a measure for the intensity of the fluctuations of the process for different timescales. Long range correlations imply that this intensity grows for all timescales. A theoretical framework for understanding fluctuation functions analytically was developed recently by Höll and Kantz [55]. It provides a direct relation to the autocorrelation function. Before, only solutions for special processes were known [53, 13]. Other properties were investigated in a number of studies both analytically and empirically [58, 130, 28, 29]. This thesis draws several new conclusions from the new theoretical understanding.

Alternative ways for obtaining fluctuation functions are detrended moving averages [8] and wavelet analysis [1]. Detrended fluctuation analysis was extended several times, for example in order to describe multifractal characteristics [67] or capture cross-correlations

[100]. These methods are not further discussed here.

Long range correlations have a strong impact on the interpretation of time averages [83]. They lead to shrinking of the effective sample size of a measured time series since the recorded events are highly dependent on each other and therefore redundant. Consequently, the error bars for the estimation of averages are much larger than for short range correlations with finite correlation time [84]. At the same time, large deviations due to perturbations decay much slower in long range correlated systems [85].

In the long history of investigations on long range correlations, they were detected in a huge number of datasets from many different fields. The existence of long range temporal correlations was found in many real world time series like in temperatures [39], biology [127], physiology [36], economics [11], and various others.

The effect is typically modeled by fractional Gaussian noise [78] in continuous time or autoregressive fractionally integrated moving average models [48] in discrete time. Both contain memory kernels and, therefore, represent the same route to long range correlations. It is known that a similar scaling of correlations can also originate from superpositions of short range correlated models [94]. It is one objective of this thesis to discuss alternative models of dynamical long range correlations, as well as emerging long range correlations due to a superposition of processes. It is also illustrated how easily short range correlations can be misinterpreted for long range correlations.

1.3 Climate variability

One of the many examples of systems where long range correlations were detected is temperature measurements around the globe. Many investigations have concentrated on the scaling of temperature autocorrelation functions. Power laws, indicating long range correlations, were found in gridded datasets covering temperatures of the whole earth with the method of detrended fluctuation analysis [39] and by investigating power spectral densities [40]. They thereby found one typical exponent for the areas of land, which are not too far from the ocean. Over the oceans, the exponents, and therefore the long memory effect, are even larger. Similar exponents were also found for station data (e.g. [84], and [91] looking at multifractal characteristics), spatially averaged temperature measurements [126], and even data produced by general circulation models [104].

The more traditional way of looking at climate variability concentrates on the search for oscillatory modes in the data. This is motivated by very pronounced phenomena like El Niño events that are known for many decades [128]. The underlying El Niño southern oscillation was described in different ways. For example, in [44] as two frequency modes, one with a 2-2.5 year period and one with 4-5 years period length. It is measured by several indices that combine measurements of sea level pressure or temperature in the equatorial Pacific area. Other oscillatory modes in the atmosphere are, for example, the Madden-Julian oscillation [76] of one and a half months and the North Atlantic oscillation [34]. In the Pacific and Atlantic multidecadal variability has also been described in a similar way [119]. The climate also oscillates on long times like for example Dansgaard-

Oeschger cycles with periods of more than a thousand years [44]. The findings are summarized in a sketch of the frequency spectrum published in [89], which illustrates that frequency modes can be found on all timescales (a similar sketch is also shown in [44]).

A different, less well known frequency mode is the 7–8 year cycle in European temperature variability. It was found using singular spectrum analysis (SSA) [99], versions of Monte Carlo SSA [96], conditional mean approaches [62], and wavelet methods [12].

Stochastic modeling of temperature data was first suggested by Hasselmann in 1976 [52]. Describing the two types of stochastic properties introduced above, lead to different types of models. The advantage of looking at (1) the scaling of fluctuations is that the methods consider the complete power of climate fluctuations. They are useful for estimating prediction errors [84, 75] and validating general circulation models [47]. It can be argued whether or not this type of dynamics is expected to be physical. For the atmosphere, the presence of such a memory kernel was motivated by modeling the ocean as infinite layers slowly interacting with each other and the atmosphere [41]. The advantage of (2) finding oscillatory modes is that they contribute to the physical understanding of the atmosphere. If they are strong they enable predictions like for example the El Niño southern oscillation [26]. They also can explain observed slowdown or speed up of global warming [119, 59].

The disadvantage of the first approach (looking at the scaling of fluctuations) is that it neglects properties that are already known. It also reduces the information to one property which is difficult to interpret. Most importantly, it is not possible to prove long range correlations with finite datasets. It has also been shown that short range correlations can easily be misinterpreted for long range correlations [82]. In many cases, short range correlations are suitable for modeling the data [81]. The disadvantage of the second approach of finding oscillatory modes is that it does not take the full power of the fluctuations into account [73]. Therefore, it never leads to a suitable data model.

In more recent work climate variability is also described with a process that combines both properties. The so-called continuum variability [61] is then treated as one of the features besides the oscillatory modes that are contained in temperature fluctuations. Recently, interest in this continuum variability is growing [74, 41]. Lovejoy [73] emphasizes its importance for understanding 'a missing quadrillion' in root mean squared error for long timescales that are not explained by the oscillatory modes. One could try to filter the known frequency modes in order to understand the continuum better.

A different approach is presented in this thesis. Here the two ways of looking at climate variability are merged by understanding the oscillations with the same method that is used to determine the scaling of autocorrelations. This method is detrended fluctuation analysis. It is shown how fluctuation functions for European station data can be understood from the superposition of oscillatory modes.

1.4 Contents and scientific contributions

In this thesis I present the results of three years of research at the Max Planck Institute for the physics of complex systems in the research group for nonlinear time series analysis. My work was supervised by Professor Holger Kantz.

This thesis is organized as follows. In the second chapter some preliminaries are presented that are essential for the understanding of the following. In chapter three, four, and five, the research results are presented and discussed. The content starts with general considerations on anomalous statistical behavior in chapter three. In chapter four and five, it concentrates on long range correlations and recorded time series that are suspected to exhibit them. The thereby developed method is later also applied to other measurements of observables in systems that exhibit complex, yet clearly not anomalous correlation patterns. In chapter five the method is extended in order to describe fluctuations on multiple timescales. All three chapters contain both new concepts as well as applications. While the applications in chapter three are all toy models, in chapter four and five they mostly include analysis of real data.

Parts of the results have been published in articles or written in the manuscripts listed in the section 'List of publications'. In the following, the contributions are introduced and linked to the table of contents of this thesis.

The article on the *Scale-invariant Green-Kubo relation for time averaged diffusivity* [C] introduces a relation that links the time averaged and the ensemble averaged version of the mean squared displacement with each other, with the velocity displacement, and the correlation function of the process. It holds for a large class of systems. Using this theory three processes are compared: a velocity renewal process, a Lévy flight, and a model of deterministic diffusion. The general relation, which was derived by my collaborator, Eli Barkai, is presented in section 3.1.1. The discussion of the different models of anomalous diffusion is mostly presented in section 3.3.3.

A more intuitive discussion on anomalous deterministic diffusion is presented in *Anomalous diffusion and the Moses effect in an aging deterministic model* [E]. The work was carried out in collaboration with Kevin Bassler and Vidushi Adlakha. Here a previously proposed framework of decomposing the root causes of anomalous diffusion was applied to the map. The results are presented in section 3.2.

My work on decomposing the effects of anomalous diffusion is currently continued in the manuscript *Moses, Noah and Joseph Effects and Infinite Densities in Coupled Lévy Processes* [J]. This is written together with Erez Aghion, Kevin Bassler and Vidushi Adlakha. The content of my contributions is included at different points in this thesis, like in section 3.3.1 and 4.1.1.

A second line of research starts with the article *Reproducing long-range correlations in global mean temperatures in simple energy balance models* [D]. Here a previously proposed energy balance model for global mean temperatures is investigated regarding the observed long range correlations in the data with detrendend fluctuation analysis. The work was done in collaboration with Marc Höll. It was shown that the observed correlations can be reproduced with a short range correlated model. The content is presented in section 4.2.

The article *Inferring characteristic timescales from the effect of autoregressive dynamics on detrended fluctuation analysis* [F] goes in more depth into understanding the properties of detrended fluctuation analysis. It also introduces the idea of fitting fluctuation functions of short range correlated models in order to obtain relaxation times and periods of noisy oscillations. These topics are discussed at several points throughout this thesis, especially in the sections 4.1.4, 4.3.1, 4.3.2, as well as 5.1.1, and 5.1.2.

In *A simple decomposition of European temperature variability capturing the variance from days to a decade* [G], an extension of the idea of fitting fluctuation functions is proposed. It enables the user to examine data with fluctuations on multiple timescales. The method is presented in section 5.1.3. In section 5.2 the application to European temperature measurements is shown. Here the method is able to reproduce a previously observed 7–8 year cycle.

The same method is applied to a fundamentally different system in *Identifying characteristic timescales in power grid frequency fluctuations with DFA* [I]. This work was done in collaboration with Mehrnaz Anvari who could link the results of the analysis to the properties of power grids. Here several characteristic timescales are quantified and two different power grids are compared. The whole analysis is presented at the end of this thesis in section 5.3.

Two additional applications of fitting fluctuation functions were first presented in the conference paper *From long memory to oscillatory modes – The potentials of detrended fluctuation analysis* [H]. They deal with sea level pressure time series. The results are presented in section 4.3.4 and section 5.1.4.

2 Preliminaries

In this chapter, the basics of some topics are introduced that will be important for the following chapters. The sections 2.1 and 2.2 deal with general and traditional concepts. Here some symbols and methods are defined that are quite commonly known. In section 2.3 and section 2.4 more advanced topics are presented. They introduce the two directions of the research that are presented in the following chapters. Therefore concepts have to be introduced that are not yet included in standard textbooks. One direction is concerned with the properties of detrended fluctuation analysis as an example of a method for detecting long range correlations. Here in section 2.3.3 a derivation for a theoretical description of the method is included. The other one deals with anomalous diffusion and its causes. Here a derivation for an infinite density is included in section 2.4.4.

2.1 Concepts and definitions

This first section is a basic introduction to the setting and notation used throughout this text.

2.1.1 The general setting

In this work, several different problems are addressed and several concepts are introduced. However, the general setting is always the same.

Complex systems are everywhere in nature. A good example is the atmosphere itself. Such a system can usually not be fully controlled and predicted analytically because on the one hand the complete state is not known due to the many degrees of freedom and on the other hand the system might exhibit chaotic behavior. Here the focus is not on trying to build a realistic model that contains the complexity of the dynamics. Instead, measurements of single observables are analyzed.

In this text, all considered datasets are one-dimensional time series $v(t)$. These datasets are naturally discrete in time. The values of the time series will usually be elements of the real numbers $\mathbb{R}$, i.e. in a continuous space. Since the real system has many degrees of freedom and it might also be driven by unknown external drivers, it is necessary to account for the resulting uncertainty. So mainly the discussed models will be stochastic. It will in many cases be assumed that the type of dynamics, i.e. the hidden real model that drives the dynamics, does not change nor depend on time.

Two types of dynamics should be distinguished. In one case the system fluctuates around one mean value. The values that an observable in a system generated by such a

process takes for different times will be denoted as $v(t)$. In the other case, the system diffuses and is not pushed back to its mean value. The corresponding time series will be called $x(t)$. A more precise definition of the difference will be given in section 2.2.

As already mentioned, measured time series must be discrete. This means that the time step duration is the smallest timescale on which information is available. It is not possible to increase the resolution without making assumptions on the process. The opposite, however, is always possible by downsampling.

In this text, where necessary, the time steps are indicated as subscript of t. This means, for example, that times with daily resolution are denoted by t_d. A series of daily measurements is therefore written as $v(t_d)$. Now, if the daily values are not of interest and a lower resolution is sufficient, $v(t_d)$ can be transformed to a time series with steps of one week $v(t_w)$ by averaging the seven days of each week

$$v(t_w) = \lfloor v(t_d) \rfloor_w \quad \text{with} \quad v(t_w) = \frac{1}{7} \sum_{j=1}^{7} v(t_d = 7t_w + j). \tag{2.1}$$

The resulting time series only has one seventh of the length of the original time series. Such transformations will be denoted as $\lfloor \rfloor_z$, where the subscript denotes the new resolution.

In order to be a true time series, the iterates of $v(t)$ should not be independent. This means that the statistical properties of $v(t)$ differ from a randomly shuffled series of the values v. So there must be some ordering in time. This is true especially in the presence of autocorrelations, which are discussed later.

Even though a large part of what will be presented in the following deals with detrended fluctuation analysis, it will usually be assumed that there is no trend in the data. In fact, the vast majority of datasets used for applications only exhibit a small trend that can be neglected, or no trend.

2.1.2 Probability distributions and ensemble averages

Now one can ask what the value of the time series $v(t)$ will be for a given time in the future. Since the considered processes are stochastic or chaotic, the answer to this question can not be a number, but rather a distribution of probabilities for different possible values. This probability distribution or density $\rho(v, t)$ is a function that depends on the continuous state variable $v \in \mathbb{R}$ and, in addition, might depend on the time $t \in \mathbb{N}$. It is normalized

$$\int_{-\infty}^{\infty} \mathrm{d}v \ \rho(v, t) = 1 \ \ \forall t. \tag{2.2}$$

The mean value of the process v at time t is then given by

$$\langle v(t) \rangle = \int_{-\infty}^{\infty} \mathrm{d}v \ v\rho(v, t) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} v_n(t). \tag{2.3}$$

This is called the ensemble average $\langle \rangle$ of the observable. The full probability density can of course not be determined with a finite number of measurements. So in practice, it is

approximated by looking at a finite set of independent realizations of $v_n(t)$. Such a set of independent realizations is called an ensemble. In this way independent and identically distributed (i.i.d.) random variables are produced since the probability distributions of different realizations of the same process are identical. The averages of i.i.d. values with finite second moment converge to the real mean value of the process according to the law of large numbers [42].

If v is a stochastic process, the mean value $\langle v(t) \rangle$ alone may not be a satisfying description. In addition, it is important how large the fluctuations around this value are, i.e. what the variance σ_v^2 of the process is. If the mean is zero, the variance is equivalent to the second moment of the process v, since

$$\sigma_v^2(t) = \langle v^2(t) \rangle - \langle v(t) \rangle^2. \tag{2.4}$$

This can be met easily if the mean value is subtracted from the time series.

The second moment is calculated in the same way as the first moment

$$\langle v^2(t) \rangle = \int_{-\infty}^{\infty} dv \; v^2 \rho(v, t) = \frac{1}{N} \sum_{n=1}^{N} v_n^2(t). \tag{2.5}$$

For diffusive processes the mean squared displacement $\langle x^2(t) \rangle$ or ensemble averaged mean squared displacement (EA MSD) is an important property of the time evolution

$$\langle x^2(t) \rangle = \frac{1}{N} \sum_{n=1}^{N} x_n^2(t). \tag{2.6}$$

Its time dependence determines the type of diffusion. An introduction to different types is given in section 2.4. Higher moments are relevant for the shape of the distribution. They are not considered in the following.

2.1.3 Time averages and ergodicity

Now, what if only one trajectory is known and no ensemble v_n can be generated? This is the realistic scenario especially in climate science where only one realization of the system exists. Averages then have to be calculated in time. The mean and the variance are usually calculated as

$$\overline{v(t)} = \frac{1}{T} \sum_{t=1}^{T} v(t), \tag{2.7}$$

$$\overline{v^2(t) - \overline{v(t)}^2} = \frac{1}{T} \sum_{t=1}^{T} v^2(t) - \overline{v(t)}^2. \tag{2.8}$$

A more interesting case are observables that show the time evolution of a diffusive system $x(t)$. This can only be captured by ensembles of series with several points in

time. If only one realization of the system exists, a pseudo-ensemble has to be created by looking at windows of the series of a certain length and averaging over the different windows. A relevant example is the time averaged mean squared displacement (TA MSD) of $x(t)$,

$$\overline{\delta^2(\Delta)} = \overline{(x(t+\Delta) - x(t))^2} = \frac{1}{T-\Delta} \sum_{t=1}^{T-\Delta} (x(t+\Delta) - x(t))^2. \tag{2.9}$$

Here the sliding window is shifted by 1 for each realization in the pseudo-ensemble. Note, that the result $\overline{\delta^2(\Delta)}$ might not only depend on the lag time Δ. If the statistics change over time, it also depends on the total time t.

The TA MSD $\overline{\delta^2(\Delta)}$ is an estimator for the mean squared displacement $\langle x^2(t) \rangle$, defined in the previous section. The property that ensures an equality is called ergodicity. For ergodic systems, the ensemble average can be replaced by the time average.

In mathematics ergodic systems are defined as systems with non-decomposable phase space, i.e. there are no ranges of values of x that a typical trajectory cannot reach. The second requirement is that the dynamics is measure preserving. Intuitively speaking, this means that the probability distribution ρ is independent of time. There are systems where the first condition is met while the second is violated. This is called weak ergodicity breaking in the literature (e.g. [18]).

2.2 Correlations in time

Since this work deals with one-dimensional time series, interactions between particles or subsystems are not discussed. Instead, the focus is on autocorrelations, which describe the dynamics of an unperturbed system. Interactions are modeled as noise. The following sections introduce the concept of correlations in time as well as some important models.

2.2.1 The correlation function and the power spectrum

The mean and the variance of a time series $v(t)$ do not carry any information about its dynamics as the ordering of the elements does not affect the result. So an additional measure is needed to describe the system. This is the autocorrelation function, defined as

$$C_{vv}(\Delta) = \frac{1}{\sigma^2} \langle v(t)v(t+\Delta) \rangle, \tag{2.10}$$

assuming stationarity, i.e. independence of t. It measures how the current state $v(t)$ depends on its past values with lag time times Δ. Similarly, it shows how future values depend on the present state. The initial value $C_{vv}(0)$ is always 1. An equivalent function which is not normalized can also be called autocovariance. In some cases, in non-stationary systems, the autocorrelation function might, in fact, depend on the absolute time t. Examples for such systems will be given later and are not considered here. For
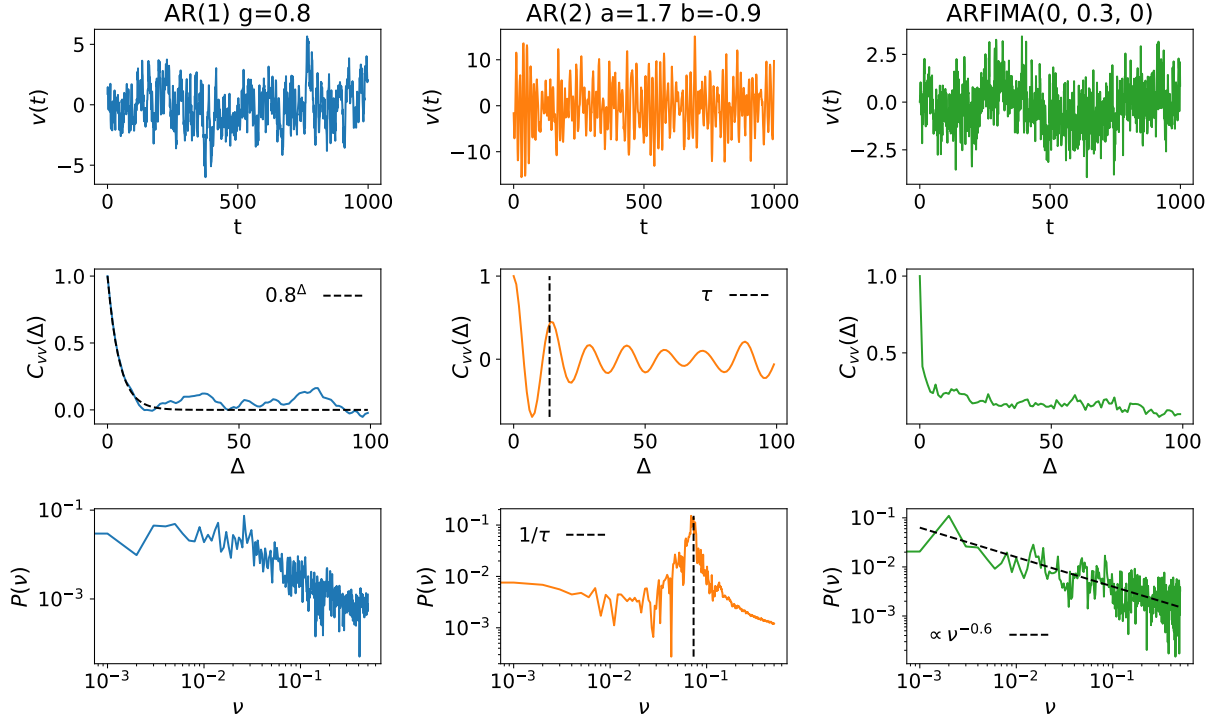
Figure 2.1

TOP: trajectories of three processes with 1000 time steps. LEFT: AR(1) (see equation (2.21)) with $g = 0.8$. CENTER: AR(2) (see equation (2.26)) with $a = 1.7$ an $b = -0.9$ exhibiting noisy oscillations. RIGHT: ARFIMA(0, d, 0) with $d = 0.3$. CENTRAL row: autocorrelation functions C_{vv} for each process. For AR(1) the exponential decay in indicated by the dashed line. For AR(2) the period τ (see equation (2.31)) is highlighted by the dashed line. BOTTOM: Power spectra calculated as $\mathcal{F}[C_{vv}]$ for each process. For AR(2) the frequency $1/\tau$ is highlighted by the dashed line. For ARFIMA both the correlation function and the power spectral density are described by a power law. Here this is highlighted by the dashed line.

---

ergodic systems with no t-dependence, the ensemble average can be replaced by a time average over t. In this case, the dependence on the lag time Δ can only be calculated for $\Delta \ll$ the total measurement time in a meaningful way. It should also be noted that the autocorrelation only reflects the linear properties of the time series [68]. An important quantity is the autocorrelation time

$$r = \sum_{\Delta=0}^{\infty} C_{vv}(\Delta). \tag{2.11}$$

If the sum $\sum_{\Delta=0}^{\infty} C_{vv}(\Delta)$ diverges the process is called long range correlated [49]. This case is further discussed in section 2.2.4.

It can now be stated more clearly what the difference is between a time series $v(t)$, which fluctuates around a mean value, and a diffusive process $x(t)$. Correlations of v,

in contrast to x, go to zero for long times, $\lim_{\Delta \to \infty} C_{vv}(\Delta) = 0$, or oscillate around zero if periodic non-stationarities are present. Explicit trends in the data can destroy this behavior since the correlation function, in this case, carries a bias. Note that for systems with long range correlations the convergence of the correlation function to zero is very slow. For systems with diverging or non-stationary variance it is useful to look at the non-normalized property $\langle v(t)v(t+\Delta) \rangle$.

A different way to describe the linear properties of the dynamics is to look at it in Fourier space. The power spectrum of a time series is a function of the frequency ν defined as [68]

$$P(\nu) = |\mathcal{F}[v](\nu)|^2, \tag{2.12}$$

where $\mathcal{F}$ is the Fourier transform

$$\mathcal{F}[v](\nu) = \frac{1}{\sqrt{T}} v(t) \exp\left(\frac{2\pi i \nu t}{T}\right). \tag{2.13}$$

According to the Wiener-Khinchin theorem, the definition of the power spectrum introduced above is equivalent to the Fourier transform of the autocorrelation function

$$P(\nu) = \mathcal{F}[C_{vv}](\nu). \tag{2.14}$$

Throughout this text the power spectrum will be calculated in this way. Examples for power spectra and autocorrelation functions are shown in figure 2.1.

The advantage of the power spectrum compared to the autocorrelation function is the effect that oscillatory dynamics have. In the power spectrum, periodic oscillations of the signal are translated to delta peaks. In the correlation function, they stay oscillations. So these regular patterns of the dynamics are clearly disentangled from noise in the power spectrum and periods are easier to measure. In the following section filtering of such oscillations will be discussed.

2.2.2 Seasonal cycles and anomalies

Spikes in the power spectrum due to regular oscillations are common, especially in atmospheric time series. The reason is the dominant role of the seasonal cycle on temperature or precipitation (depending on the latitude). These periodic patterns are a challenge when analyzing the fluctuations of the data [33]. Therefore in most studies, a simple trick is used in order to filter them. This is done by calculating the so-called anomalies (see e.g. [39, 84]). Since the same trick is applied several times throughout this text is deserves to, at least, be briefly introduced at this point.

It works for pure delta peaks in the power spectrum no matter if it describes a daily cycle in data with one minute resolution or a yearly cycle in daily data. So for all periods of $\tau = n\Delta t$ where n is a positive integer and Δt is the step length of the time series. For seasonal cycles, the method can only be used if the values for the 29th of February in leap years are removed.

In the first step the mean cycle $S(t)$ is calculated by averaging the values that appear at a specific phase value of the oscillation, e.g. on the same calendar day

$$S(\Delta) = \frac{1}{N} \sum_{n=1}^{N} v(\tau n + \Delta). \tag{2.15}$$

Then the cycle is subtracted from the time series, which yields the anomalies $\tilde{v}(t)$

$$\tilde{v}(t) = v(t) - S(\Delta) \quad \text{with } \Delta = t \bmod \tau, \tag{2.16}$$

using the modulo operator. In the resulting signal, the regular oscillations are suppressed.

For some other time series, like sea level pressure recordings, the variance exhibits clear seasonality [G]. This is not a linear property and will not be analyzed in depth here, but data models can be improved if here like for the oscillations of the mean, anomalies are calculated and the cycles are added to the model. The cycle $S'(\Delta)$ can be calculated analogous to $S(\Delta)$. The detrended time series $\tilde{\tilde{v}}(t)$ can be calculated as

$$\tilde{\tilde{v}}(t) = \frac{\tilde{v}(t)}{S'(\Delta)} = \frac{v(t) - S(\Delta)}{S'(\Delta)}. \tag{2.17}$$

The anomalies can then be described by stationary model.

2.2.3 Autoregressive models

Autoregressive (AR) processes are a popular tool which provides basic data models for time series. They are linear filters of Gaussian white noise $\epsilon(t)$. In this section a summary of the most important properties is presented. A full introduction can be found in [21].

Autoregressive processes of order p without trends are defined by an iteration relation

$$X(t) = \sum_{k=1}^{p} \phi_k X(t - k) + \epsilon(t), \tag{2.18}$$

where ϕ_k are the p AR-parameters. Equivalently to equation (2.18) the process can also be defined using the backshift operator B with $B^k x(t) = x(t - k)$ ($k \in \mathbb{N}$). It then reads

$$\Phi(B)X(t) := \left(1 - \sum_{k=1}^{p} \phi_k B^k\right) X(t) = \epsilon(t), \tag{2.19}$$

where $\Phi(B)$ was defined as a short notation for the expression in the brackets. The equation can be rewritten as

$$\prod_{k=1}^{p} (1 - g_k B) \, x(t) = \epsilon(t), \tag{2.20}$$

where the $g_k^{-1} \in \mathbb{C}$ are the roots of the characteristic equation $\phi(B) = 0$.

In this text only the two most simple versions of autoregressive processes will appear. They are now described in a bit more detail in order to clarify the correlation structure they describe. The AR(1) model (autoregressive model of order one),

$$X(t) = gX(t-1) + \epsilon(t), \tag{2.21}$$

describes linear response to Gaussian white noise. The parameter in chosen $|g| < 1$ for stability. Here only the case $g > 0$ is discussed. The correlation function can be calculated from the difference equation

$$C(\Delta) = gC(\Delta - 1), \tag{2.22}$$

which has the solution [42]

$$C(\Delta) = g^\Delta = e^{-\Delta|\log g|}. \tag{2.23}$$

This is analogous to the corresponding differential equation in continuous time. Note that the reasoning is independent of the exact shape of the increment distribution ϵ. The autocorrelations are the same if the noise is not Gaussian. The relaxation time is equivalent to the correlation time r. It reads

$$r = \frac{-1}{\log(g)}. \tag{2.24}$$

The standard deviation of the process is given by

$$\sigma_X^2 = \frac{\sigma_\epsilon^2}{1 - g^2}. \tag{2.25}$$

The AR(2) model is the simplest model to go beyond linear response. It is defined by

$$X(t) = aX(t-1) + bX(t-2) + \epsilon(t), \tag{2.26}$$

where $\epsilon(t)$ is again Gaussian white noise and $a \in [-2, 2]$, and $b \in [-1, 1]$ are the AR-parameters. The variance of the process is

$$\sigma_X^2 = \frac{(1-b)\sigma_\epsilon^2}{(1+b)(1-a-b)(1+a-b)}, \tag{2.27}$$

depending on the noise variance σ_ϵ^2.

Equation (2.26) can be rewritten as

$$(1 - g_1 B)(1 - g_2 B)X(t) = \epsilon(t), \tag{2.28}$$

In this notation the roots g_1, $g_2 \in \mathbb{C}$ replace the AR-parameters. The correlation function of AR(2) is given by [21]

$$C(t) = h_1 g_1^t + h_2 g_2^t, \tag{2.29}$$

where constants h_1 and h_2 are calculated from $C(t = 1) = a/(1 - b)$ and $C(t = 0) = 1$. Accordingly, the correlation time can be calculated to be

$$r = -\frac{h_1}{\log|g_1|} - \frac{h_2}{\log|g_2|}. \tag{2.30}$$

Again, note that the correlation function is not a property of the distribution of the process and Gaussianity of the noise is not a necessary condition for equation (2.29).

If the roots are complex, the dynamics describes a noisy oscillation. The period can be determined to be

$$\tau = \frac{2\pi}{\arctan(\mathcal{I}[g_2]/\mathcal{R}[g_2])}, \tag{2.31}$$

with the imaginary part $\mathcal{I}[g_2]$ and the real part $\mathcal{R}[g_2]$ of g_2.

2.2.4 ARFIMA

The autoregressive-fractionally-integrated-moving-average (ARFIMA(p, d, q)) process, sometimes also called FARIMA or fractionally integrated ARMA (see [21] for an introduction), is a model for long memory in time series. It is a generalization of the previously introduced AR(p) process by a moving average of some order q, and a fractional integration with fraction $d \in [-1/2, 1/2]$. The moving average will be neglected in this text. Therefore the defining equation capturing autoregression and fractional integration is

$$\Phi(B)(1 - B)^d X(t) = \epsilon(t). \tag{2.32}$$

Here B is again the backshift operator and $\epsilon(t)$ is Gaussian white noise. The operator $\Phi(B)$ generates autoregressive dynamics and was defined in the previous section. In the case of ARFIMA(1, d, 0) it reads $(1 - gB)$.

The operator $(1-B)^d$ describes the fractional integration, which leads to long memory. For $d > -1$ it can be written as the binomial expansion

$$(1 - B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)}{\Gamma(k + 1)\Gamma(-d)} B^k =: \sum_{k=0}^{\infty} \pi_k B^k, \tag{2.33}$$

where $\Gamma(k)$ is the well known gamma-function

$$\Gamma(k) = \int_0^\infty t^{k-1} \exp(-t)\mathrm{d}t \sim \sqrt{2\pi} \exp(-k + 1)(k - 1)^{k-1/2}. \tag{2.34}$$

The approximate solution in equation (2.34) is the so-called Stirling formula, which is valid for large k. Alternatively the π_k can be defined via the recursion

$$\pi_k = \frac{k - 1 - d}{k} \pi_{k-1}. \tag{2.35}$$

The best way to understand the fractional integration is by looking at the dynamics of a process ARFIMA(0, d, 0), where all other effects are absent. According to equation

(2.32) the autoregression and the fractional integration can be separated. Therefore, ARFIMA(p, d, 0) is equivalent to an autoregressive process with fractionally integrated noise as input instead of white noise $\epsilon(t)$. The same separation is possible for moving averages [21]. So, to understand the full process, it is sufficient to understand each of the three effects individually.

Fractionally integrated white noise is written as

$$(1 - B)^d X(t) = \epsilon(t). \tag{2.36}$$

Its variance can be calculated as

$$\sigma_X^2 = \sigma_\epsilon^2 \frac{\Gamma(1 - 2d)}{\Gamma(1 - d)^2}, \tag{2.37}$$

depending on the variance of the white noise ϵ. The autocorrelation function can be calculated, too. It reads

$$C(\Delta) = \frac{\Gamma(\Delta + d)\Gamma(1 - d)}{\Gamma(\Delta - d + 1)\Gamma(d)}. \tag{2.38}$$

For long lag times Δ the autocorrelation function can be approximated by the Stirling formula (2.34), to be

$$C(\Delta) \approx \Delta^{2d-1} \frac{\Gamma(1 - d)}{\Gamma(d)}, \tag{2.39}$$

which scales like a power law with exponent $2d - 1$. Similarly the power spectrum $P(\nu)$ for small ν scales like $P(\nu) \sim \nu^{-2d}$. For all $d > 0$ the correlation time r diverges

$$\sum_{t=0}^{\infty} C(t) \to \infty. \tag{2.40}$$

This property is called long range correlations or long memory. It leads to several interesting effects like slower decay of large deviations [85], larger error in mean estimation [84], and anomalous diffusion [31].

2.3 Detrended fluctuation analysis

By direct calculation of the correlation function, it is difficult to find out whether a process is long range correlated or not. This is because the correlation function is a monotonously decreasing function in this situation. Therefore, the tail for long times is close to zero, where statistical deviations are larger than the actual value. A popular alternative method is detrended fluctuation analysis developed in 1994 [97]. Its analytical description is very recent [55].

2.3.1 Fluctuation analysis and time averages

It would be beneficial to have an increasing function with a power law behavior, which can be fitted in a log-log-plot. Such a function is given by the fluctuation function F of fluctuation analysis (FA) [65].

In order to obtain an increasing function, it is first useful to look at the discrete integral of $v(t)$ instead of the values themselves. The cumulative sum (or integral) $x(t)$, sometimes also called the profile of the time series, is

$$x(t) = \sum_{n=1}^{t} v(n).\qquad(2.41)$$

with $x(0) := 0$. The squared fluctuation function quantifies the variance of x in segments of different lengths s. The segments are chosen in such a way that they do not overlap. Therefore, for each length s, there are $\max\{N \in \mathbb{N} | N \leq (T/s)\}$ different segments. Here T is the length of the time series and the expression indicates the maximum integer number to fulfill the condition. The squared fluctuation function F^2 is a function of this segment length. It is defined as

$$F^2(s) = \frac{1}{N} \sum_{n=1}^{N} \Big(x(ns) - x((n-1)s) \Big)^2.\qquad(2.42)$$

Usually one is interested in the asymptotic scaling of this function

$$F(s) \propto s^{\alpha} \quad \text{for} \quad s \to \infty.\qquad(2.43)$$

It is obtained by looking at the log-log plot and fitting a straight line. If the time series $v(t)$ is uncorrelated, $F(s) \propto s^{1/2}$. The same scaling with $1/2$ can still be observed asymptotically if the process is short range correlated. For long range correlated systems the scaling is always $\alpha > 1/2$. For ARFIMA(0, d, 0), for example, the scaling is $\alpha = d + 1/2$ [29]. If the analysis yields $\alpha < 1/2$, this implies anti-persistence and 'intermediate memory' [29] in the system. Even though the correlation time is finite, such processes exhibit a Hurst effect.

The fluctuation function can be compared to the TA MSD defined in section 2.1.3

$$\overline{\delta^2(\Delta)} = \frac{1}{T - \Delta} \sum_{t=1}^{T-\Delta} (x(t + \Delta) - x(t))^2.\qquad(2.44)$$

Both, in fact, are time averages of some quantity $\overline{(x(t + \Delta) - x(t))^2}$ with $\Delta = s$. This quantity is calculated for several windows, which are averaged. The difference is that the window for the TA MSD is shifted by one for each time step, while for FA only non-overlapping windows are considered. Accordingly the window is shifted by s. The results are not necessarily identical. However, both methods estimate the same quantity. The difference is how many windows are taken in order to estimate it. DFA uses less windows and is therefore faster. For both methods the windows are identically distributed in time and only the resolution is different. So long time properties are the same, at least up to some factor. The scaling α is a property that can be calculated in both ways.
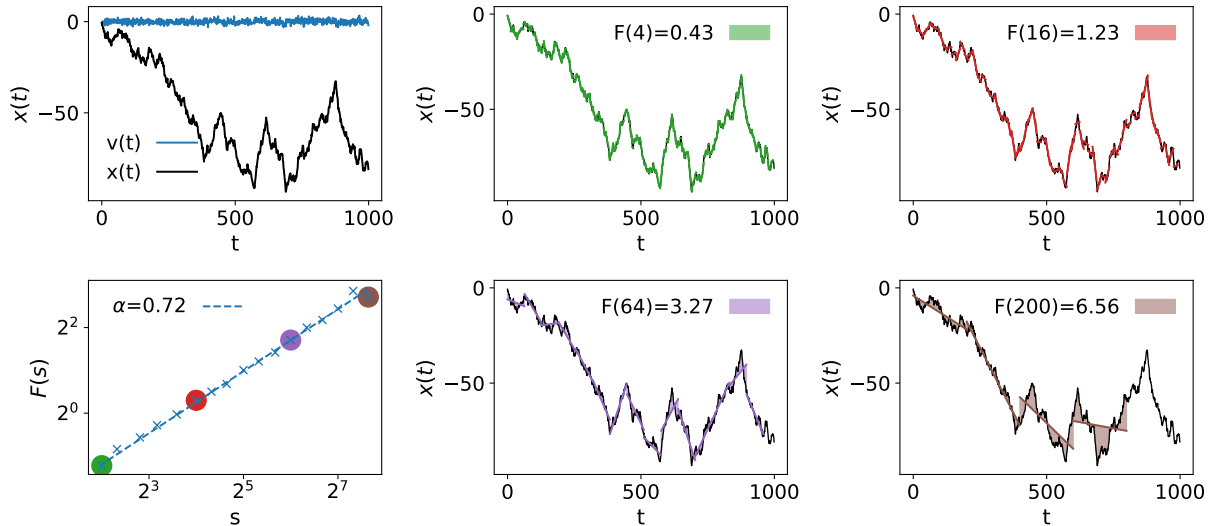
Figure 2.2
TOP LEFT: The trajectory of the time series $v(t)$ plotted together with the sum $x(t)$. CEN-TRAL and RIGHT column: calculation of $F(s)$ for different s in DFA1; the fits in the segments are plotted with colors, the area between the fit and the time series $x(t)$, which represents the amplitude of the fluctuations is filled with the same color. BOTTOM LEFT: the obtained values for $F(s)$ plotted in a log-log plot and fitted by a polynomial. The obtained scaling is $\alpha = 0.72$. The points for which the fits are shown in one of the other panels are highlighted by large circles.

2.3.2 Definition of the detrended fluctuation function

Detrended fluctuation analysis is an algorithm that is used in order to calculate a fluctuation function $F^2(s)$, which is similar to the one introduced in the previous section. However, it remains useful for asymmetric data, diffusive dynamics, and data with polynomial trends. The detrending can be implemented with a chosen order depending on a parameter q. The corresponding version of DFA is then called DFAq.

The algorithm will be presented in the following. Similar to FA, the first step of DFA is to calculate the cumulative sum $x(t)$, where

$$x(t) = \sum_{n=1}^{t} v(n). \tag{2.45}$$

Then the time series $x(t)$ is divided into N non-overlapping segments of length s. To each of these segments a polynomial $p_{n,s}(t)$ of order q is fitted and subtracted. The fluctuations in the segment are quantified by the variance $V^2(\nu, s)$ of the difference time

series. It is calculated as

$$V^2(n, s) = \frac{1}{s} \sum_{t=1}^{s} (x(t + (n-1)s) - p_{n,s}(t))^2. \tag{2.46}$$

The order q is the detrending order. DFAq is able to remove polynomial trends of order $q-1$. So non-stationarities of lower order (or very slow fluctuations on timescales beyond the measurement time) can be neglected.

DFA does not consider non-stationarity of the fluctuations. The information contained in the variability for segments of length s is reduced to one quantity, which is the value of fluctuation function $F^2(s)$. It is calculated as the mean value of $V^2(n, s)$ in all N segments

$$F^2(s) = \frac{1}{N} \sum_{n=1}^{N} V^2(n, s). \tag{2.47}$$

Often the sum goes to $2N$ because the procedure described above is applied not only to $x(t)$, but in addition also to a version of with reversed ordering $x(-t)$. This is done in [66] because the procedure for each s only takes into account $N \times s$ values and the last values in the time series are ignored. By adding values of $V^2(n, s)$ obtained from a time series with reversed ordering all values enter the calculation. The result of DFA is therefore invariant for reversed time. It does not matter, whether the variance grows or decays over time.

Picking the maximal value for s is a trade-off between limitations of DFA and the ability to get any meaningful fit from a small dataset. Usually, it is recommended to take at most 1/4th of the length of the time series [65].

The detrending filters the slowly varying dynamics and enables the user to see the short time fluctuations. Obviously, there is no clear classification of what a trend is and it cannot be distinguished from slow fluctuations. Also, DFA becomes increasingly unstable for higher detrending orders. So it should be avoided to use a higher detrending order than necessary for the target of the investigation. In fact, most applications, shown in this text, only require DFA1.

2.3.3 Theory of DFA

As shown by Höll and Kantz [55] the fluctuation function $F(s)$ of DFA can be calculated analytically as a transformation of the stationary autocorrelation function $C(\Delta)$. This result was not the first analytical approach for understanding DFA [121, 53], however, it was the first which showed a general and direct relation that can be used for calculations. It is essential for the idea of fitting characteristic timescales with DFA, which will be introduced in chapter 4. Therefore, the derivation, presented in [56], will be recalled in this section.

The stationary autocorrelation function $C(\Delta)$ is not necessarily equivalent to the autocorrelation function of the time series. It is assumed that the correlation function of the signal can be decomposed into a non-stationary component, which is due to a

trend that is filtered by DFA and the correlation function of the stationary noise. The latter is considered here. The theory is based on one crucial assumption, namely, the statistical properties of the fluctuations have to be stable throughout the recorded time series, i.e. variances are statistically equal for each segment of a certain length. In the notation of equation (2.46), this means

$$\langle V^2(n,s)\rangle = \langle V^2(1,s)\rangle \forall \nu. \tag{2.48}$$

If the condition (2.48) is fulfilled, the expected values of the fluctuation function simplify to

$$\langle F^2(s)\rangle = \frac{1}{N} N \langle V^2(1,s)\rangle = \langle V^2(1,s)\rangle = \frac{1}{s}\sum_{t=1}^{s}\langle (x(t)-p_{1,s}(t))^2\rangle. \tag{2.49}$$

Now one has to find an analytical expression for the fitting polynomial $p_{1,s}(t)$. In [56] it was shown, that it can be written as

$$p_{1,s}(t) = \sum_{k=1}^{s} v(k) P_s(t,k), \tag{2.50}$$

where the $P_s(t,k)$ are calculated by

$$P_s(t,k) = \frac{1}{\det(S)} \sum_{i=1}^{q+1} t^i \sum_{j=1}^{q+1} \mathrm{adj}(S)_{i,j} \sum_{r=k}^{s} r^j. \tag{2.51}$$

S is a $q+1 \times q+1$ matrix, $\det(S)$ is the determinant and $\mathrm{adj}(S)$ represents the adjugate matrix. The elements of S are defined as

$$S_{i,j} = \sum_{k=1}^{s} k^{i+j-2}. \tag{2.52}$$

The profile $x(t)$ can be written in a similar form to the results of the fit $p_{1,s}(t)$. Using the Heaviside function $\Theta(t)$, which is zero for all negative t and one for $t \leq 0$, it reads

$$x(t) = \sum_{k=1}^{s} v(k)\Theta(t-k). \tag{2.53}$$

Accordingly, by inserting equation (2.50) and (2.53) in (2.49), one obtains

$$\langle F^2(s)\rangle = \frac{1}{s}\sum_{t=1}^{s}\left\langle \left(\sum_{k=1}^{s} v(k)(\Theta(t-k)-P_s(t,k))\right)^2 \right\rangle. \tag{2.54}$$

The last equation can be written in a more compact form with the symbol

$$\mathcal{P}_s(t,k) = \Theta(t-k) - P_s(t,k) \tag{2.55}$$

for the expression in the brackets.

The square of the sums in equation (2.54) can now be applied. Rearranging the terms yields

$$\langle F^2(s) \rangle = \frac{1}{s} \sum_{k=1}^{s} \left(\langle v^2(k) \rangle \sum_{t=1}^{s} \mathcal{P}_s^2(t, k) \right)$$
$$+ \frac{2}{s} \sum_{\Delta=1}^{s-1} \sum_{k=1}^{s-\Delta} \left(\langle v(k)v(k+\Delta) \rangle \sum_{t=1}^{s} \mathcal{P}_s(t, k)\mathcal{P}_s(t, k+\Delta) \right). \quad (2.56)$$

This expression depends only on two properties of the values $v(t)$ of the time series. Firstly, on the variance $\sigma^2 = \langle v^2(t) \rangle$ and, secondly, on the autocorrelation function $C(\Delta) = \langle v(k)v(k+\Delta) \rangle / \sigma^2$ for stationary v. The fluctuation function is therefore determined by 2-point correlations and the second moment. Higher moments do not appear. Accordingly, the fluctuation function can be expressed as

$$\langle F^2(s) \rangle = \sigma^2 \frac{1}{s} \sum_{k=1}^{s} \left(\sum_{t=1}^{s} \mathcal{P}_s^2(t, k) \right) + \frac{2}{s} \sum_{\Delta=1}^{s-1} \sum_{k=1}^{s-\Delta} \left(C(\Delta) \sum_{t=1}^{s} \mathcal{P}_s(t, k)\mathcal{P}_s(t, k+\Delta) \right). \quad (2.57)$$

Finally, a kernel L can be defined which includes the sums over the $\mathcal{P}$ in both terms

$$L(\Delta, s) = \frac{1}{s} \sum_{k=1}^{s-\Delta} \sum_{t=1}^{s} \mathcal{P}_s(t, k)\mathcal{P}_s(t, k+\Delta). \quad (2.58)$$

By using this kernel, the fluctuation function can be written in the final form

$$\langle F^2(s) \rangle = \sigma^2 \left(L(0, s) + 2 \sum_{t=1}^{s-1} C(t)L(t, s) \right). \quad (2.59)$$

Thus the relationship between the fluctuation function and the correlation function is established. On the one hand, the fluctuation function depends on two properties of the time series, namely $\sigma^2 = \sigma_v^2$ and $C(t) = C_{vv}(t)$. On the other hand, it depends on the kernel L and, therefore, on the detrending order q. To be precise, one should write L_q. So the fluctuation function F depends on both, $v(t)$ and q. For clarity, it will later be denoted by $F_v(s)$ with a subscript in order to show which data is described.

The kernel L_q was calculated in [56] for the lowest orders of detrending, which are the most relevant ones. The case of DFA0 is equivalent to FA, which was presented in section (2.3.1). Here the kernel reads

$$L_0(\Delta, s) = \frac{1}{6s^2}(-\Delta^3 + 3s\Delta^2 + \Delta(-3s^2 + 1) + s^3 - s). \quad (2.60)$$

For $q = 1$, with linear fitting polynomials, the kernel L_1 in equation (2.59) is

$$L_1(\Delta, s) = \frac{1}{30(s^4 - s^2)}[3\Delta^5 - 5(4s^2 - 1)\Delta^3 + 30(s^3 - s)\Delta^2$$
$$- (15s^4 - 35s^2 + 8)\Delta + 2(s^5 - 5s^3 + 4s)]. \quad (2.61)$$

The two upper mentioned cases are different in the way they respond to diffusing dynamics like Brownian motion. While DFA0 always yields values smaller or equal to one, DFA1 in the case on Brownian motion yields $\alpha = 1.5$ [53]. A more detailed discussion on this issue will be presented in section 4.1.1. The higher orders only lead to higher filtering of slow dynamics. For $q = 2$ the kernel L_2 is

$$
\begin{aligned}
L_2(\Delta, s) = & \frac{1}{70(s^6 - 5s^4 + 4s^2)}[-10\Delta^7 + (42s^2 - 28)\Delta^5 - 35(3s^4 - 9s^2 + 2)\Delta^3 \\
& + 105(s^5 - 5s^3 + 4s)\Delta^2 + (-35s^6 + 280s^4 - 497s^2 + 108)\Delta \\
& + 3(s^7 - 14s^5 + 49s^3 - 36s)].
\end{aligned}
\tag{2.62}
$$

Note that with higher q the expressions for L_q become longer and therefore less handy. However, they can always be calculated using equation (2.58) and they all can be written in a similar polynomial form. Higher orders are relevant if the data exhibits a clear trend like, e.g., climate change in atmospheric data. In this text most applications use DFA1.

At the end of this section, there should be one remark on the relation of DFA to similar methods like the wavelet analysis presented in [1] and detrended moving averages [8]. Both have the same aim as DFA, namely, they measure the scaling of the correlation function for long times. At the same time, they are able to detrend the time series in a way that slow dynamics with polynomial shape can be neglected. In fact, Höll and Kantz [57] found out that they can be formulated analytically in a similar way, namely, equation (2.59) holds for all these detrending methods. The difference is the kernel L, which depends on the method. Apparently, many different kernels could be used for the same purpose.

2.4 Anomalous statistical properties

Long range correlations, which were discussed in the previous sections, are one anomalous statistical property. It is a property that violates the premises of the central limit theorem and therefore leads to anomalous diffusion. In the following, this theorem, as well as the term anomalous diffusion, will be introduced in sections 2.4.1 and 2.4.2. Then, two more anomalous properties, other than long range correlations, are introduced in sections 2.4.3 and 2.4.4. Both deal with the probability distribution of the values v.

2.4.1 Central limit theorem

The central limit theorem is an important result in the field of statistics. It is the reason for the dominant role of the Gaussian distribution in many research fields [21]. There are different versions and generalizations of this theorem.

In its basic form the central limit theorem states that the sum

$$
x(t) = \sum_{s=1}^{t} v(s),
\tag{2.63}
$$

of independent and identically distributed (i.i.d.) random variables $v(t)$ with finite variance σ_v^2 is distributed like a Gaussian random variable. The width of this distribution scales with time like $\sqrt{t}$. So a rescaled version of $x(t)$ obtains a stable Gaussian distribution

$$\rho\left(\frac{x(t) - \langle x(t)\rangle}{\sqrt{t}}\right) = \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(\frac{(x(t) - \langle x(t)\rangle)^2}{2\sigma_v^2}\right). \tag{2.64}$$

Note that there are three requirements for the central limit theorem to hold. These are independence of the increments $v(t)$, existence of an identical probability distribution $\rho(v)$ for all increments (i.e. stationarity of $v(t)$), and existence of the second moment and, therefore, the variance σ_v^2 of this distribution.

These are the traditional assumptions for the central limit theorem, however, they are not the weakest possible requirements for a Gaussian distribution to emerge. Suppose, for example, the values $v(t)$ are not independent, but themselves sums of Δ independent random variables $w(t)$ given as

$$v(t) = \sum_{n=t}^{t+\Delta} w(n). \tag{2.65}$$

Then in the definition of $x(t)$ the values v could be replaced by the values w and the theorem would hold for these. However, since the variance σ_v^2 appears in equation (2.64) the parameters of the Gaussian distribution would have to be adjusted. Nevertheless, for sufficiently fast decaying correlations of the increments $v(t)$, some version of the central limit theorem is always expected [42], at least for long times t beyond the correlation time.

2.4.2 The Hurst effect

If all conditions of the central limit theorem are met, the mean squared displacement $\langle x^2(t)\rangle$ of the integrated process x scales linearly with time

$$\langle x^2(t)\rangle \sim 2Dt \quad \text{for } t \to \infty, \tag{2.66}$$

with the diffusion constant D. This behavior is called normal diffusion. The process, which is the sum of independent and identically distributed random variables with finite variance, is called Brownian motion. If one condition of the central limit theorem is violated, the mean squared displacement might still follow a scaling law, i.e. exhibit power law dependence on time, at least for long times. This phenomenon is called anomalous diffusion and the exponent is equal to twice the so-called Hurst exponent H:

$$\langle x^2(t)\rangle \sim 2Dt^{2H} \quad \text{for } t \to \infty. \tag{2.67}$$

There are several mechanisms, that might lead to this behavior with $H \neq 1/2$. They will be introduced in this text.

One arises if the increments are not independent but instead long range correlated. This is the case for example in the ARFIMA model introduced in section 2.2.4. The
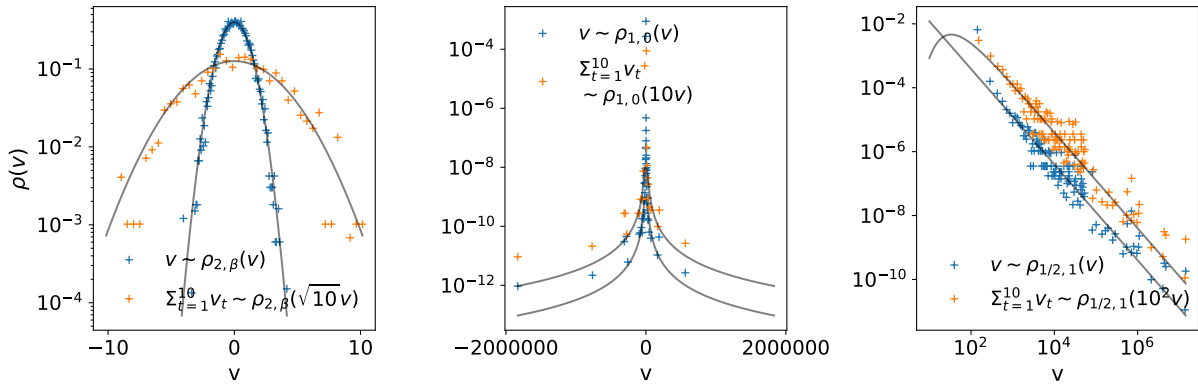
Figure 2.3

Generalized central limit theorems for three special cases. The initial distribution of an ensemble is compared to the distribution of the sum of $N = 10$ independent increments. LEFT: the Gaussian distribution $\rho_{2,\beta}(v)$ reproduces the standard central limit theorem with a scaling of $\sqrt{N}$. CENTER: the Cauchy distribution $\rho_{1,0}(v)$ scales like N. RIGHT: the one-sided Lévy-Smirnov distribution $\rho_{1/2,1}(v)$ yields a scaling of N^2.

corresponding anomalously diffusing process can be described by a scale-invariant Green-Kubo relation [31]. A second mechanism, which is easy to understand, is present in scaled Brownian motion [63], where the increment distribution $\rho(v)$ is not identical for all $v(t)$, but explicitly depends on time. Both routes to anomalous diffusion will be discussed in section 3.1.3.

In the following two sections, two more situations are presented, in which the increment distribution is not identical or its variance diverges, yet the process exhibits scaling.

2.4.3 Lévy stable distributions

One condition for the Central Limit theorem to hold is that the variance of the probability distribution of the values v has to be finite. In fact, this condition is only necessary to ensure a Gaussian shape of the sum x. In general, the sum of independent identically distributed random variables always converges to some well-defined distribution if it is normalized correctly. This is ensured by the generalized central limit theorem. It is explained in this section without proof following the presentation in [25].

All distributions that can be realized as the limit distribution of a superposition of i.i.d. variables are called Lévy stable distribution. In general, they can be defined by their characteristic function, i.e. their Fourier transform

$$\mathcal{F}[\rho_{\gamma,\beta}(x,\mu,\sigma)] = \rho_{\gamma,\beta}(k,\mu,\sigma) = \exp\left[i\mu k - \sigma^\gamma |k|^\gamma \left(1 - i\beta \frac{k}{|k|} w(k,\gamma)\right)\right], \qquad (2.68)$$

where the expression for $w(k,\gamma)$ depends on the value of γ. It is defined in the interval

$0 \leq \gamma \leq 2$ in the following way:

$$w(k, \gamma) = \begin{cases} -\frac{2}{\pi} \ln |k| & \text{if} \quad \gamma = 1 \\ \tan \frac{\pi\gamma}{2} & \text{if} \quad \gamma \neq 1 \end{cases}. \tag{2.69}$$

The index γ is most important for the asymptotic decay of the distribution. It also determines the scaling of the sums of i.i.d. random variables. Stable distributions are conserved with rescaling

$$\rho_{\gamma,\beta} \left(\frac{\sum_{n=1}^{t} v(n)}{t^{1/\gamma}} \right) = \rho_{\gamma,\beta} \left(\frac{x(t)}{t^{1/\gamma}} \right) \tag{2.70}$$

In general, the characteristic function, and therefore also the probability distribution, depends on four parameters β, γ, μ and σ. The skewness parameter β defines the asymmetry of the distribution. It is always in the interval $-1 \leq \beta \leq 1$. The distribution is symmetric only if $\beta = 0$. $\rho_{\gamma,\beta}$ is the mirror image of $\rho_{\gamma,-\beta}$. If $\beta = \pm 1$, the probability distribution is extremal. In this case, the distribution is one-sided if $\gamma < 1$ at the same time.

The parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ simply shift and scale the probability distribution and do not play a role for integrability of the probability density function. In fact, one can always rescale according to

$$\rho_{\gamma,\beta}(x, \mu, \sigma) = \frac{1}{\sigma} \rho_{\gamma,\beta} \left(\frac{x - \mu}{\sigma}, 0, 1 \right). \tag{2.71}$$

For the purpose of this section it is sufficient to look at the properties of $\rho_{\gamma,\beta}(x) = \rho_{\gamma,\beta}(x, 0, 1)$.

The crucial property of the stable distributions is the behavior in the tails for $x \to \infty$. Processes, that follow a generalized central limit theorem instead of the traditional one, deviate from others in their tails. Due to more extreme events, their distributions have infinite variances. The asymptotic behavior of the stable distributions is determined by the parameter γ. In the pre-factor, the Gamma function $\Gamma(x)$ appears. The complete expression is

$$\rho_{\gamma,\beta} \overset{|x|\to\infty}{\sim} \frac{1}{\pi} \sin \left(\frac{\pi\gamma}{2} \right) \Gamma(1 + \alpha) |x|^{-(1+\gamma)}. \tag{2.72}$$

The second moment diverges for $\gamma < 2$.

There are only three cases in which the probability density ρ can be written down explicitly and not only be defined by the characteristic function. Here all three cases are shown in figure 2.3. As they represent different types of stable distributions, they are a useful sample of the set of stable distributions. One of these cases is $\alpha = 2$. Here the generalized central limit theorem coincides with the traditional central limit theorem. The probability distribution is Gaussian

$$\rho_{2,\beta} = \frac{1}{\sqrt{4\pi}} \exp \left(-\frac{x^2}{4} \right). \tag{2.73}$$

The parameter β is unnecessary, since $\tan(\pi) = 0$ in equation (2.69).

The second case is a symmetric distribution with $\beta = 0$ and $\gamma = 1$. It is called a Cauchy distribution and reads

$$\rho_{1,0}(x) = \frac{1}{\pi(1 + x^2)}. \tag{2.74}$$

Finally the third case is the asymmetric Lévy-Smirnov distribution for $\gamma = 1/2$ and $\beta = 1$. It reads

$$\rho_{1/2,1} = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-3/2} \exp\left(-\frac{1}{2x}\right) & x \geq 0 \\ 0 & x < 0 \end{cases}. \tag{2.75}$$

Solutions with other parameters can be expressed via the Fox H - functions.

The three special cases introduced above are reproduced numerically in figure 2.3. The scaling with N is tested by comparing the initial distribution to the distribution of the sum of 10 i.i.d. random variables.

2.4.4 Aging and infinite densities

Distributions with power law tails might not only be present as i.i.d. noise. Instead, it might also be implemented as waiting times between two events. Here obviously only one-sided distributions are plausible.

An example of a discrete deterministic system that exhibits such waiting times is the Pomeau-Manneville map. This system will be discussed in more detail in section 3.2. In this section, the phenomena of aging and infinite invariant densities are explained with this map as a relevant example. The effects are, however, much more general. They are present in many stochastic processes [102, 2]. In its classical version, the map [101] reads

$$v(t + 1) = D(v(t)) = v(t) + av^z(t) \bmod 1, \tag{2.76}$$

with the modulo operator that subtracts the largest integer $< v$ and positive parameters a and z. If v is close to zero, the system makes very small steps away from the origin, thus the dynamics is regular and slow. It can be interpreted as a waiting time. When the trajectory exceeds unity, it gets reinjected. Two successive reinjections can be assumed independent because the following reinjection is very sensitive to the exact starting point of the regular period.

The waiting time distribution is proportional to $w(\tau) \sim \tau^{z/(z-1)}$ [43]. For $z > 2$ the mean waiting time is infinite, which leads to a phenomenon called aging. In this section only the case $z > 2$ is discussed, since in this regime the interesting phenomena emerge. This means that the probability distribution is not stationary. The effect can be understood by looking at the waiting time distribution from a starting point $t > 0$. The longer the system is already running, the higher is the probability that this starting point lies in an extremely long waiting interval. This so-called aged waiting time distribution $w_0(\tau, t)$ therefore depends on the time t. It is known from the literature of aged continuous time random walks and presented in [15] as

$$w_0(\tau, t) \sim \frac{\sin(\pi\alpha)}{\pi} \frac{t^{\frac{1}{z-1}}}{\tau^{\frac{1}{z-1}}(t + \tau)}. \tag{2.77}$$
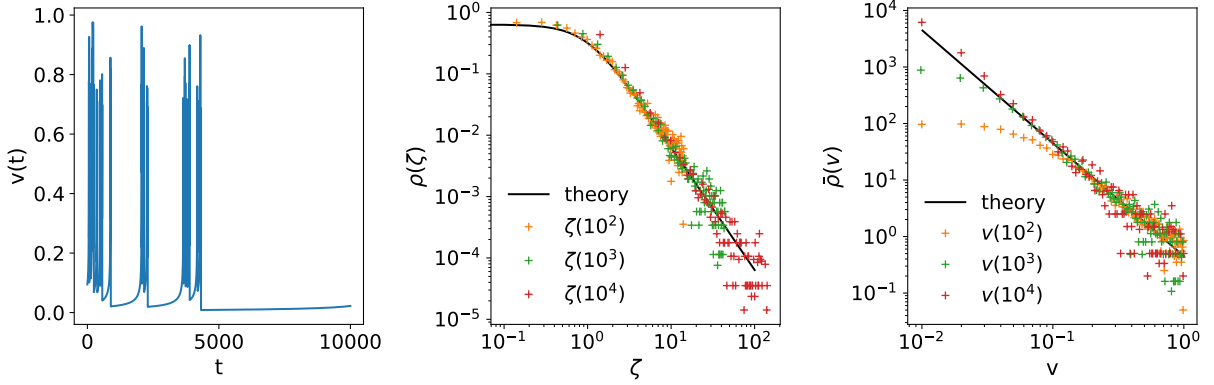
Figure 2.4

LEFT: A trajectory of the Pomeau-Manneville map (2.76) with $z = 3$. CENTER: the corresponding rescaled density (2.85) of an ensemble of such systems for different running times starting from a uniform distribution. All distributions collapse to one theory curve, described by equation (2.86). RIGHT: distributions of the same values, but this time scaled according to equation (2.88). For long times the curves approach the theoretical function (2.89).

For long waiting times the distribution is proportional to $\tau^{1/(z-1)}$ instead of $\tau^{z/(z-1)}$.

In the following, the corresponding behavior of the probability distribution will be derived. The presentation follows the calculation of Akimoto and Barkai [3]. The crucial property of the map is its behavior close to $v \approx 0$ since it determines the length of the waiting periods. One can start with defining $(0 : \chi(\Delta))$ as the interval that is not left before time Δ. Then the probability of being still inside the interval $(0 : \chi(\Delta))$ at time t is equal to the probability of waiting times being larger than Δ for an aged system with t.

In the first step, it is useful to have an approximation for the paths $\chi(\Delta)$ during one waiting time. It is derived in [122] (section 3) to be

$$\chi(\Delta) \approx \Big(a(z-1)\Delta \Big)^{\frac{1}{1-z}}. \tag{2.78}$$

This behavior is due to

$$\Delta \sim \int_{\chi(\Delta)}^{\chi_{max}} \mathrm{d}v\ \frac{1}{v - D(v)} \sim \int_{\chi(\Delta)}^{\chi_{max}} \mathrm{d}v\ \frac{1}{av^{z-1}} = \frac{\chi(\Delta)^{1-z} - \chi_{max}^{1-z}}{a(z-1)} \sim \frac{\chi(\Delta)^{1-z}}{a(z-1)}, \tag{2.79}$$

where χ_{max} is the largest value to be reached in the waiting period. The first step is proven in [122] in more detail for a more general class of functions D.

Now consider the probability of v still being located the interval $(0 : \chi(\Delta))$ closer to zero than $\chi(\Delta)$. It can be related to the probability of long waiting times, which also corresponds to v being located close to zero. The relation,

$$\int_0^{\chi(\Delta)} \mathrm{d}x \rho(v, t) \sim \int_\Delta^\infty \mathrm{d}\tau w_0(\tau, t), \tag{2.80}$$

follows from equation (2.78). By differentiating both sides and plugging in equations (2.77) and (2.78), this yields

$$\rho\left((a\Delta(z-1))^{\frac{1}{1-z}}, t\right)\frac{1}{\Delta^{\frac{z}{z-1}}} \sim \frac{1}{\Delta^{\frac{1}{z-1}}(t+\Delta)}. \tag{2.81}$$

Now the argument in the distribution can be transformed back to v. By replacing the parameter Δ with v one finds that

$$\rho(v,t) = A\frac{1}{1+a(z-1)v^{z-1}t}, \tag{2.82}$$

where A is a normalization constant due to ρ being a density. It can be calculated and plugged in, which yields the final probability distribution

$$\rho(v,t) \approx \frac{\sin(\pi/(z-1))}{\pi/(z-1)}(t(z-1))^{\frac{1}{z-1}}\frac{1}{1+a(z-1)v^{z-1}t}. \tag{2.83}$$

The result describes the behavior of the density close to $v \approx 0$. It depends explicitly on the time t, for which the system is running (aging time of the system).

The good news is, that the time dependence follows a simple power law. This means the density (2.83) can be regarded as a stationary distribution of a rescaled variable

$$\zeta = |v|\left(a(z-1)t\right)^{\frac{1}{1-z}}. \tag{2.84}$$

Then the rescaled density is obtained by taking

$$\varrho(\zeta) = \frac{1}{(at(z-1))^{1/(z-1)}}\rho\left((at(z-1))^{1/(z-1)}\zeta, t\right). \tag{2.85}$$

The explicit representation reads

$$\varrho(\zeta) = \frac{\sin(\pi/(z-1))}{\pi/(z-1)}a^{-\frac{z-2}{z-1}}\frac{1}{1+\zeta^{z-1}a^{2-z}}. \tag{2.86}$$

The last equation gives a form for the rescaled density that no longer depends on the time t. It was first derived by Dynkin [35] in the context of renewal theory, while Thaler [123] established the connection to the underlying transformations. The central panel of figure 2.4 shows how the rescaled density collapses to one curve. It describes the behavior of the density close to zero. Depending on the parameter z, the moments of the distribution might all or partly diverge.

These moments can be understood by looking at a second way to rescale equation (2.83) and obtain an invariant expression, namely the infinite invariant density. This is done by not rescaling the variable v but the distribution $\rho(v)$ instead. Of course, the resulting function is no longer normalizable and can therefore not be interpreted as a physical density. However, it is a very useful quantity for calculating moments of the distribution (2.83). Considering the long aged system $t \to \infty$, the '+1' in the

denominator can be neglected. In this case the remaining dependence on the time t scales with $t^{-\frac{z-2}{z-1}}$. The infinite invariant density $\bar{\rho}$ is therefore

$$\bar{\rho}(v) = \lim_{t \to \infty} t^{\frac{z-2}{z-1}} \rho(v,t) \approx \frac{\sin(\pi/(z-1))}{\pi} a^{-\frac{z-2}{z-1}} (z-1)^{\frac{1}{z-1}} v^{1-z}. \tag{2.87}$$

In general infinite densities [102] are probability measures, which are not normalizable. They are known from infinite ergodic theory. Generally, they are defined as functions $\hat{\rho}(v/t^{\beta})$ with some exponent β. The infinite covariant density is the limit

$$\hat{\rho}(v/t^{\beta}) = c \lim_{t \to \infty} t^{\gamma} \rho(x,t), \tag{2.88}$$

with some proportionality constant $c > 0$. The Lévy central limit theorem is a subclass of the infinite covariant densities for $\beta = \gamma$.

For the special case of $\beta = 0$ the so called infinite invariant density can be defined as

$$\hat{\rho}(x) = c \lim_{t \to \infty} t^{\gamma} \rho(x,t). \tag{2.89}$$

This is the case for the Pomeau-Manneville map for $z > 2$.

The derivation of the scaled density in this section shows a mechanism of how the probability distribution can become non-stationary. In this way, the premise if identically distributed increments in the central limit theorem is violated. The previous section 2.4.3 showed implications of fat tailed distributions that violate the premise of finite variances of the increments. In the following chapter, the effects of these violations will be formalized and discussed in an understandable framework. An application for the theory of the Pomeau-Manneville map is presented in section 3.2.

3 Identifying the causes of anomalous diffusion

This chapter presents and discusses a framework of decomposing the causes that lead to anomalous diffusion. Such a framework is relevant when analyzing data for both interpretation and model selection. In the first part, the decomposition is derived and explained. In the second part, it is applied to a model of deterministic anomalous diffusion generated by the Pomeau-Manneville map. This model pushes the theory to its limits because its dynamical origin of anomalous diffusion differs from the paradigmatic models introduced so far. The third part summarizes further observations of types of dynamics that generate different traces in the framework of decomposing the Hurst effect.

It can be concluded that the interpretation of the three effects remains meaningful for all considered cases and that the proposed scaling relation can be proven for a large class of systems.

3.1 The decomposition

Anomalous diffusion is a consequence of violations of the premises of the central limit theorem. Therefore it can be decomposed into different effects of violating different conditions. In order for the central limit theorem to hold for a process $x(t) = \sum_{s=1}^{t} v(s)$, the increments $v(t)$ have to be

- independent (or at least become independent after some time, then the central limit theorem will hold for coarse-grained data),

- identically distributed, and

- have a distribution with a finite second moment.

The effects which result from a violation of any of these three conditions have been named the Joseph effect, the Moses effect, and the Noah effect [79, 27]. In the following such violations are considered that still lead to scaling in time, i.e. the systems are self-affine [65]. This means that the second moment of the distribution (the MSD) grows in time with a power law. The upper mentioned effects can then be quantified by the Joseph exponent J, the latent exponent L, and the Moses exponent M. They can be related to the Hurst exponent of anomalous diffusion.

In the first section 3.1.1 a derivation of the scale-invariant Green-Kubo relation for time averaged diffusivity is presented. In the second section 3.1.2 it is shown how

this calculation proves the scaling relation and therefore the decomposition of the Hurst exponent into the three effects. In the third section 3.1.3 applications to the paradigmatic models of the Joseph, Noah and Moses effects are shown.

3.1.1 Scale-invariant Green-Kubo relation

In order to understand the meaning of different measures of anomalous diffusion in this section, a relation between the scaling of the ensemble averaged time averaged mean squared displacement (EA TA MSD), the scaling of the correlation function, of the velocity displacement, and the ensemble averaged mean squared displacement (EA MSD) is derived. The calculation was originally performed by Eli Barkai and is published in [C].

Consider diffusive processes $x(t) = \int_0^t ds \ v(s)$ with zero mean. Note that here the process is continuous in time. The result in the end will turn out to be valid for discrete time series as well. In the long time limit the EA MSD exhibits scaling with an exponent of $2H > 0$

$$\langle x^2(t) \rangle = 2Dt^{2H}, \tag{3.1}$$

thus the process is an example for anomalous diffusion if $H \neq 1/2$.

Now assume the velocity correlations are of the form

$$\langle v(t)v(t+\tau) \rangle = Ct^{2H-2}\phi\left(\frac{\tau}{t}\right). \tag{3.2}$$

The class of processes which follow this equation is very general for processes that exhibit anomalous scaling. Most processes in this class, however, are not stationary but exhibit aging due to the dependence of the correlation function on the total measurement time t. Unlike in stationary processes, the autocorrelation function is not identical for all times t. Therefore it makes a difference if the ensemble average $\langle \rangle$ is taken or a time average.

The Green-Kubo relation connects the diffusion constant D to the velocity autocorrelation function for processes that exhibit normal diffusion [71]. In [31] it was shown that this relation can be generalized for a process of the type introduced above. It can be calculated from the following equation for the EA MSD

$$\langle x^2(t) \rangle = 2 \int_0^t dt_2 \int_0^{t_2} dt_1 \langle v(t_2)v(t_1) \rangle. \tag{3.3}$$

This holds because on the one hand

$$\langle x^2(t) \rangle = \left\langle \int_0^t dt_2 \frac{d}{dt_2}x^2(t_2) \right\rangle = \int_0^t dt_2 \ \langle 2x(t_2)v(t_2) \rangle, \tag{3.4}$$

and on the other hand

$$2\int_0^t dt_2 \int_0^{t_2} dt_1 \langle v(t_1)v(t_2) \rangle = \int_0^t dt_2 \ \left\langle 2\int_0^{t_2} dt_1 v(t_2)v(t_1) \right\rangle = \int_0^t dt_2 \ \langle 2x(t_2)v(t_2) \rangle, \tag{3.5}$$

where the symmetry of the autocovariance function was used.

This derivation is only valid if the integrals are finite. Therefore some additional assumptions on the process have to be made. $\phi(q)$ is positive valued and the correlation functions decay monotonically. More explicitly $\phi(q)$ scales like [31]

$$\begin{array}{llll} \phi(q) < c_1 q^{-\delta_1} & \text{with} & 2 - 2H \leq \delta_1 < 1 & q \to 0 \\ \phi(q) < c_u q^{-\delta_u} & \text{with} & \delta_u > 1 - 2H & q \to \infty. \end{array} \tag{3.6}$$

In the processes considered below the variance of the velocity is asymptotically

$$\langle v^2(t) \rangle \sim a\mathcal{C}t^\beta, \tag{3.7}$$

with the exponent β

$$-1 \leq \beta < 2H - 1. \tag{3.8}$$

The system is energy conserving only if $\beta = 0$.

Now continuity demands due to equation (3.6) and equation (3.7) that for small q

$$\delta_1 = 2 - 2H + \beta. \tag{3.9}$$

This relation gives the scaling of the correlation function for small-q. It can be described by the two exponents β and H which are both measurable using equations (3.7) and (3.1). This gives the first relation between the Hurst exponent, the variance of the increments and their correlation function. In the following, this relation is used to find the properties of the time averaged mean squared displacement (TA MSD) of $x(t)$.

The TA MSD in continuous time is defined as

$$\overline{\delta^2} = \frac{1}{t} - \Delta \int_0^{t-\Delta} \mathrm{d}t_0 \left(x(t_0 + \Delta) - x(t_0) \right)^2, \tag{3.10}$$

similar to the definition (2.9). Here t is the measurement time up to which the path $x(s)$ is recorded. $\Delta \ll t$ is the lag time within the window for which the squared displacement is calculated. If the system is ergodic, the scaling of $\overline{\delta^2}$ with Δ is the same as the scaling of the ensemble averaged quantity $\langle x^2(t) \rangle$ with t, i.e. two times the Hurst exponent. In this case, the result of $\overline{\delta^2}$ does not depend on t. For other processes, which will be discussed here, this is not the case. Therefore $\overline{\delta^2}$ has to be investigated separately. In many cases, the time average may remain a random variable and not converge to a number but only to a distribution [86], even in the long time limit. In fact, the scaling seems to be very stable in many cases even if the time average is not defined. An example is shown in section 3.3.3. Nevertheless, for an analytical calculation, it is necessary here to consider the EA TA MSD $\langle \overline{\delta^2} \rangle$, which is well defined and converges.

Using equation (3.10) and $t \gg \Delta$, it is possible to split the integral into the short time behavior and the long time behavior

$$\langle \overline{\delta^2} \rangle \simeq \frac{1}{t} \int_0^{K\Delta} \mathrm{d}t_0 \left\langle \left(x(t_0 + \Delta) - x(t_0) \right)^2 \right\rangle + \frac{1}{t} \int_{K\Delta}^t \mathrm{d}t_0 \left\langle \left(x(t_0 + \Delta) - x(t_0) \right)^2 \right\rangle. \tag{3.11}$$

Here K is some large number separating the two ranges. Therefore it should satisfy $\Delta \ll K\Delta \ll t$. The second integral grows if the measurement time t is increased. So in the limit $t \to \infty$, the first integral can be neglected. In order to evaluate the second integral the EA MSD between the times t_0 and $t_0 + \Delta$ has to be found. Concentrating on this value one can write

$$\left\langle \left(x(t_0 + \Delta) - x(t_0) \right)^2 \right\rangle = 2 \int_0^\Delta dt_2 \int_0^{t_2} dt_1 \, \langle v(t_1 + t_0)v(t_2 + t_0) \rangle. \tag{3.12}$$

Accordingly $\tau = t_2 - t_1$ and $t = t_1 + t_0$ when equation (3.12) is compared to the definition of the autocorrelation function function (3.2)

$$\left\langle \left(x(t_0 + \Delta) - x(t_0) \right)^2 \right\rangle = 2\mathcal{C} \int_0^\Delta dt_2 \int_0^{t_2} dt_1 \, (t_1 + t_0)^{\nu - 2} \phi \left(\frac{t_2 - t_1}{t_1 + t_0} \right). \tag{3.13}$$

If the aging time is long $t_0 \gg \Delta$ the correlation function is dominated by the small q behavior described in equation (3.9), i.e. δ_1 defined in equation (3.6). In this limit the integral reads

$$\left\langle \left(x(t_0 + \Delta) - x(t_0) \right)^2 \right\rangle \sim 2\mathcal{C} \int_0^\Delta dt_2 \int_0^{t_2} dt_1 \, a \, (t_1 + t_0)^\beta \, (t_2 - t_1)^{-2 + 2H - \beta}, \tag{3.14}$$

which can be calculated. One finds for $t_0 \ll t_1$

$$\left\langle \left(x(t_0 + \Delta) - x(t_0) \right)^2 \right\rangle \sim 2\frac{a\mathcal{C}_{EA}}{(2H - \beta - 1)(2H - \beta)}(t_0)^\beta \Delta^{2H - \beta}. \tag{3.15}$$

The main equation of this section can now be derived by inserting equation (3.15) in equation (3.11) and performing the integration

$$\langle \overline{\delta^2} \rangle \sim \frac{2c_1\mathcal{C}}{(\beta + 1)(2H - \beta - 1)(2H - \beta)}t^\beta \Delta^{2H - \beta}. \tag{3.16}$$

This describes the EA TA MSD in terms of the Hurst exponent, the scaling of the velocity displacement β, and the autocorrelation function. Note that the dependence on the correlation function is not on its full shape $\phi(q)$, but only on the small q behavior. This is different from the scaling Green-Kubo relation for the EA MSD [31].

In general, the time average (3.16) behaves very differently from the ensemble average (3.1). This is the case if the system exhibits weak ergodicity breaking. Only in the case of energy conservation $\beta = 0$ the scaling is the same in both functions and the time averaged mean squared displacement does not depend on the measurement time t. If not only the system is energy conserving $\beta = 0$, but also $H = 1/2$ the presented theory is not valid any more, according to equation (3.8). This is the standard case where ensemble average and time average are equivalent, the system exhibits normal diffusion, and the transport coefficient is determined by the standard Green-Kubo relation. No theory of anomalous diffusion is needed here.

3.1.2 Joseph, Noah and Moses

The Josef exponent J can be defined by the scaling of the EA TA MSD with the lag time Δ

$$\langle \overline{\delta^2} \rangle \sim \Delta^{2J}. \tag{3.17}$$

Comparing this definition with equation (3.16) yields a relation between the exponents

$$2J = 2H - \beta. \tag{3.18}$$

Here the Hurst exponent H is defined by the scaling of the EA MSD $\langle x^2(t) \rangle \sim t^{2H}$. It represents the width of the distribution in an anomalously diffusing process. The exponent β is defined by the scaling of the velocity displacement $\langle v^2(t) \rangle \sim t^\beta$. It represents the change in the energy of the system. As β is normally around zero, the direct numerical calculation is unstable. Performance can be improved by looking at the sum over the squares, thus the function Z can be defined [27]

$$Z(t) = \sum_{s=1}^{t} v^2(s). \tag{3.19}$$

The ensemble mean or ensemble median of $Z(t)$ scales like $\langle Z(t) \rangle \sim t^{\beta+1}$. In [27] it was recommended to use the median rather than the mean because it yields more stable results. A detailed discussion of when both might yield results that are not equivalent is yet to be done for this setting.

According to [27], a deviation from $\beta = 0$ can be explained by two effects. One of them is the Noah effect, which describes fat tailed distributions of the increments (velocities). It is known to cause ergodicity breaking and anomalous diffusion, as i.i.d. processes lead to different Lévy-stable distributions. This scaling of fat-tailed distributions was introduced in section 2.4.3. The Noah effect leads to divergence of $\langle v^2(t) \rangle$. The more increments $v(t_i)$ are considered, the more probable is the occurrence of a very big value, which leads to anomalous scaling in equation (3.19).

The second effect is the Moses effect. It is present in systems with non-stationary velocity distributions. While for these processes $\langle v^2(t) \rangle$ does not diverge, it explicitly depends on t. The standard example for this behavior is Scaled Brownian Motion [63]. In this thesis, it will always be modeled as a discrete process, where the increments explicitly depend on time.

In order to distinguish both effects, it is useful to define a second function Y as the sum over the absolute values of the increments [27]

$$Y(t) = \sum_{s=1}^{t} |v(s)|. \tag{3.20}$$

While the scaling of this function is not affected by a Noah effect, as long as the first moment of the distribution is finite, it is affected by explicit non-stationarities in the

same way as Z. Therefore the Moses exponent M is defined such that $M = \frac{1}{2}$ for stationary processes. It can be calculated by

$$\langle Y(t) \rangle \sim t^{M+1/2}, \tag{3.21}$$

again using the ensemble average or the median. Accordingly, the Latent exponent L, corresponding to the Noah effect, is the remaining contribution to β. It can be defined by the scaling of the function $Z(t)$

$$\langle Z(t) \rangle \sim t^{2L+2M-1}. \tag{3.22}$$

It is $L = \frac{1}{2}$ for velocity distributions with finite second moment.

Now the scaling relation (3.18) can be rewritten with respect to the four exponents H, J, L, and M. It is the same that was proposed in [27]

$$H = J + L + M - 1. \tag{3.23}$$

Of course, the Joseph exponent can in reality not be calculated by the time-continuous sliding window average used here since data is always discrete. It can, however, be approximated by repeatedly sliding the window by one time step. An alternative way of calculating the Joseph exponent is the algorithm of DFA0. It was explained in section 2.3.1 that the window averaging used for DFA0 is equivalent to the sliding window. Since all three approaches sample the windows of the real time series from a uniform distribution in time, their scaling is equivalent.

The Hurst exponent H is defined as the scaling of the EA MSD. However, it is also possible to numerically calculate it by looking at different measures of the width of the distribution of $x(t)$. In [27, E] it is calculated as the scaling of the difference between the first and the third quantile.

3.1.3 The exponents in models

In a given process one or more of the three effects can be present. For each exponent, J, L, and M, there is a popular paradigmatic model where all other exponents are $\frac{1}{2}$. Results for these models as well as combinations were derived in [27] (see supplementary material).

For the Joseph effect, it is fractional Brownian motion. In the discrete version this can be produced by adding up increments generated by an ARFIMA(0, d, 0) model

$$x(t) = \sum_{s=1}^{t} v(s), \quad \text{with} \quad (1 - B)^{J-1/2} v(s) = \epsilon(s). \tag{3.24}$$

Here the long range correlations originate from a memory kernel. Increments are calculated by taking into account all past increments. A trajectory, as well as a numerical calculation for all four exponents in this model, is shown in the left panels of figure 3.1. Here the Joseph exponent is calculated using DFA0 defined in section 2.3.1. Accordingly, the exponent seems to be always $J \leq 1$.
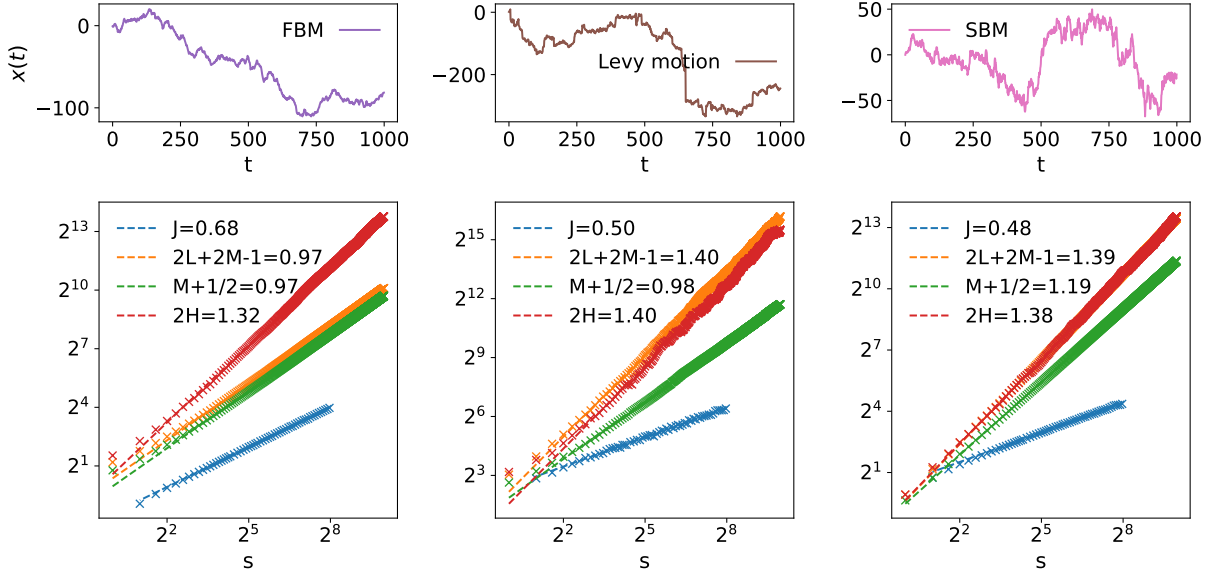
Figure 3.1

The paradigmatic models for the three effects of anomalous diffusion. All three considered processes have the same Hurst exponent $H = 0.7$. TOP: trajectory of the process with length of 1000 steps, BOTTOM: calculations of all four exponents for the same process. LEFT: fractional Brownian motion with a theoretical exponent of $J = 0.7$. The numerical calculation yields a result close to the true value. It is performed using 1000 realizations of the process. CENTER: Lévy motion with a theoretical exponent of $L = 0.7$. It is reproduced with an ensemble of 1000 by taking the median for each time point s. RIGHT: scaled Brownian motion with a theoretical exponent of $M = 0.7$. The numerical calculation yields a result close to the true value. Again it is performed using 1000 realizations of the process.

The paradigmatic model for the Noah effect is the sum over uncorrelated increments, where the probability distribution of the increments has fat tails. This means the tail asymptotically decays like a power law. As explained in section 2.4.3, for such a sum of increments with identical fat tailed distributions, a generalized central limit theorem (Lévy central limit theorem) exists. This ensures convergence to a stable distribution with some scaling in time described by a power law

$$\rho(x(t)) = t^L \rho\left(x(t)/t^L\right). \tag{3.25}$$

Such a model,

$$x(t) = \sum_{s=1}^{t} v(s), \quad \text{with} \quad \langle v(s)v(s+\Delta)\rangle = \delta(\Delta), \quad \text{and} \quad \lim_{v\to\infty} \rho(|v|) \propto |v|^{-(L+1)}. \tag{3.26}$$

can therefore be called Lévy motion. For this model, the trajectory, as well as the measurement of the exponents, are displayed in the central panels of figure 3.1. The calculation is more difficult compared to the other effects due to weak ergodicity breaking. So the median is taken instead of the mean for the numerical calculation in the figure.

Finally, the Moses effect can be understood from scaled Brownian motion. For a discrete version here one simply has to introduce a time-dependent factor to regular Brownian motion. So the model reads as follows:

$$x(t) = \sum_{s=1}^{t} v(s), \quad \text{with} \quad v(s) = s^{M-1/2}\epsilon(s), \tag{3.27}$$

with Gaussian white noise ϵ. The explicit time dependence affects the first moment of the distribution as well as the second moment. However, L is always $1/2$. Trajectory and numerical calculation of the exponents can be found in the right panels of figure 3.1.

3.2 The Pomeau-Manneville map

The Pomeau-Manneville map [101] is the standard example for dynamical intermittency. It is one-dimensional and bounded. Here a symmetric version of the map is used for generating a velocity time series $v(t)$. In this section, first, the map is introduced and then the exponents H, L, M and J of the corresponding diffusive process are calculated and interpreted. The results presented in this section were published in [E]. This work was done in collaboration with Kevin Bassler and Vidushi Adlakha.

3.2.1 The map

Consider a one-dimensional discrete-time process $x(t)$, and the corresponding velocity time series $v(t)$, which can also be considered as the increments of the process

$$x(t) = \sum_{s=1}^{t} v(s). \tag{3.28}$$

The probability distribution of v should be symmetric around zero. Here the velocities are iterates of the modified Pomeau-Manneville map

$$v(t+1) = D(v(t)) = \begin{cases} -4v(t) + 3 & \text{if} \quad 0.5 < v(t) \leq 1.0 \\ v(t)\left(1 + |2v(t)|^{z-1}\right) & \text{if} \quad |v(t)| \leq 0.5 \\ -4v(t) - 3 & \text{if} \quad -1 \leq v(t) < -0.5 \end{cases} \tag{3.29}$$

This is a symmetric version of the map [101] introduced in section 2.4.4. Maps of Pomeau-Manneville type have been studied extensively in the past [70]. For a large range of choices for the parameter z, it exhibits anomalous statistical behavior including aging [15]. This property manifests itself in the infinite invariant density and scaled density shown in section 2.4.4. A consequence of the non-stationarity is weak ergodicity breaking [18]. Spatially extended versions of the map, as well as the integrated time series $x(t)$, diffuse anomalously [43, 93].

The map is plotted in figure 3.2 in the left panel. In the central panel a trajectory $x(t)$, generated by the map, is shown. The right panel shows the autocorrelation function of
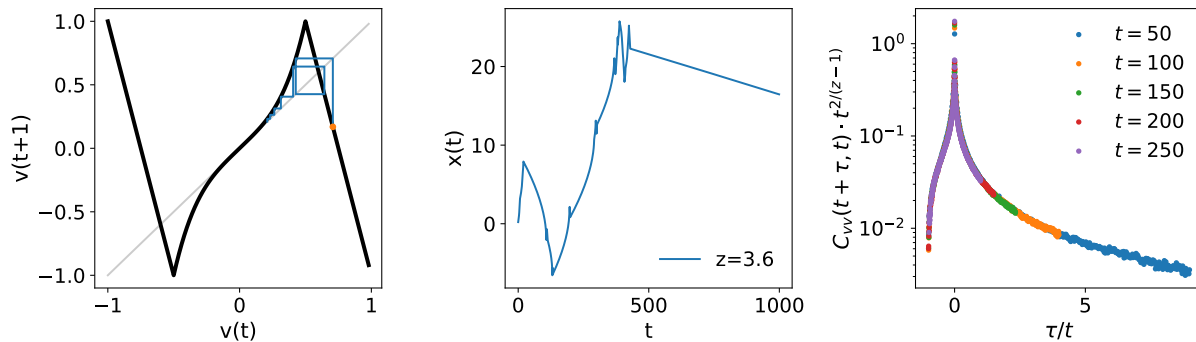
Figure 3.2

The symmetric Pomeau-Manneville map for $z = 3.6$. LEFT: The map with a trajectory indicated as a cobweb plot. CENTER: A trajectory of the integrated process $x(t)$ with 1000 time steps. RIGHT: Scaled correlation function of $v(t)$. The x-axis shows the fraction $q = \tau/t$ while the y-axis shows the autocorrelation rescaled by $t^{2-\nu}$. Thus the graph shows the function Φ from equation (3.2), multiplied by $\mathcal{C}$. The result for different times (potted as points in different colors) nicely coincide, which shows that the Pomeau-Manneville map for $z > 2.5$ is a system of the type described by equation (3.2).

the increments $v(t)$ for different times. By rescaling, the functions at different times can be matched together. This numerically proves that the Pomeau-Manneville map is a system of the type described in section 3.1.1, at least for $z > 2.5$ [C].

The dynamics generated by the map is intermittent. The system stays very long close to $v \approx 0$ and exhibits regular repulsion from this fixed point. Once the trajectory reaches the outer region $|v| > 1/2$, it undergoes a chaotic burst and eventually gets reinjected to the regular region around $v \approx 0$. In the following section, the stochastic properties of this dynamics shall be introduced. These properties are crucial for understanding the diffusion of $x(t)$. The map can, therefore, be understood by drawing parallels to stochastic processes. The idea of generating diffusion deterministically can be regarded as an anomalous version of what was presented in [17]. The possibility of generating this type of dynamics in a deterministic system raises confidence in the applicability to the real world. It is even possible to have the same properties in a system in continuous time. Such a system was introduced in [B]. This continuous-time version is a nonlinear periodically driven oscillator. The same behavior of the infinite density as for the map was detected for this system. It also exhibits the same scaling.

The initial velocity $v(1)$ is chosen randomly from a uniform distribution in the interval $[-1, 1]$. This is very close to the initial distribution of the rescaled density (2.86) for $t = 1$. The results for the rescaled density as well as the infinite invariant density were derived in section 2.4.4. In order to calculate the functions $Y(t)$ and $Z(t)$ only the density of the absolute value is needed. Therefore, the results from section 2.4.4 are still valid for the symmetric map.

3.2.2 Stochastic approximation and the Hurst effect

In order to find a corresponding stochastic description of the dynamics, it is first necessary to analytically understand the regular behavior of the map during the waiting times. Similarly to section 2.4.4 for the derivation of the infinite density, the dynamics around the marginally stable fixed point at zero the system can be approximated by a continuous differential equation [5]

$$\frac{\mathrm{d}\hat{v}}{\mathrm{d}t} = \hat{v} \, |2\hat{v}|^{z-1} \text{ for } \hat{v} \to 0, \tag{3.30}$$

where $\hat{v}$ is the continuous approximation of v. The absolute value of v is slowly but monotonously increasing. Once v escapes the regular region a chaotic burst occurs before it gets reinjected. Typically the chaotic motion in the outer region is very short. Its contributions to x are therefore small. The reinjection points are uncorrelated with each other due to the chaotic burst in between. They determine the paths and therefore the waiting time inside the inner region. Therefore, the dynamics could be approximated by randomly generated waiting times, which imply a certain dynamics until the next reinjection. Such behavior is equivalent to stochastic renewal processes with well defined waiting time distributions [112].

The waiting time distribution $w(\tau)$ can be calculated by integrating equation (3.30). It is solved by [6]

$$\hat{v}(t) = \chi_0 \left(1 - \frac{t}{\tau} \right)^{\frac{1}{1-z}}, \tag{3.31}$$

with the waiting time τ depending on the reinjection points χ_0

$$\tau = \frac{1}{z-1}(2\chi_0)^{1-z}. \tag{3.32}$$

By looking at a fixed maximum value of $v = 1$ in equation (3.31), it is easy to see that τ correctly describes the waiting time in the regular interval up to some factor [88]. Now assuming a uniform distribution of reinjection points χ_0 (which is justified [43, 6]), equation (3.32) yields a transformation of the probability distribution that determines the waiting time distribution $w(\tau)$ as

$$w(\tau) \propto \tau^{\frac{z}{1-z}}. \tag{3.33}$$

For $z \leq 2$ this expression has a well defined mean. In this case there is no Hurst effect. In the following, the case of $z > 2$ with infinite mean waiting time is discussed. During each waiting period, the process $x(t)$ performs many small steps v in the same direction away from the origin. The total displacement χ during one waiting period starting at the reinjection time $t = t_r$ is

$$\chi = x(t_r + \tau) - x(t_r) = \int_{t_r}^{t_r+\tau} \mathrm{d}s \, \hat{v}(s). \tag{3.34}$$

Using the equations (3.30,3.33), it can be calculated to be [C]

$$\chi \sim \frac{1}{2}(z-1)^{\frac{2-z}{1-z}} \frac{1}{z-2} \tau^{\frac{2-z}{1-z}}. \tag{3.35}$$

This establishes a relation between χ and the waiting times τ. In the theory of stochastic renewal processes this relation is usually given by the joint probability density function for χ and τ [87]. Here such a relation can be written as

$$P(\chi,\tau) \propto \tau^{\frac{z}{1-z}} \frac{1}{2} \left[\delta(\chi - \tau^{\frac{2-z}{1-z}}) + \delta(\chi + \tau^{\frac{2-z}{1-z}}) \right], \tag{3.36}$$

with the Dirac δ-function.

It was shown in [93] that the value of H can be calculated from the theory of continuous time random walks, which are stochastic renewal processes. The crucial information is given by equation (3.36). The fact that the process generated by the map is discrete, in contrast to renewal processes, does not matter here. The same is not necessarily true for the other exponents J, L and M. This issue will be discussed in section 3.3.3.

For continuous time random walks the calculation can be performed in Fourier-Laplace space [5]. There the mean squared displacement is equal to the second derivative of the spatial distribution of the process. Results are known for joint probability distributions of the form of equation (3.36). They were calculated for Lévy flights which perform jumps instead of escapes with finite velocity [5]. H is also known for Lévy walks [7], where the velocity during a waiting period is constant. Since the results in both cases are the same, and in fact are independent of the exact path, they also apply to the process $x(t)$ generated by the Pomeau-Manneville map [B, C, 93]. The Hurst exponent reads

$$H = \begin{cases} 0.5 & \text{if} \quad z < 2 \\ 0.5/(z-1) & \text{if} \quad 2 < z < \frac{5}{2} \\ (z-2)(z-1) & \text{if} \quad \frac{5}{2} < z \end{cases}. \tag{3.37}$$

The system exhibits normal diffusion for $z < 2$. For $2 < z < 3$ it is subdiffusive and for $z > 3$ it is superdiffusive. At the crossover points at $z = 2$ and $z = 2.5$ the values of H change continuously. However, logarithmic corrections appear for these parameter values [5].

3.2.3 Calculating the exponents

The exponents L and M can be calculated using the results presented in section 2.4.4. As the exponents are defined via the probability distributions of the absolute value $|v|$ of v, it makes no difference whether the traditional Pomeau-Manneville map is considered or the symmetric version. The rescaled density can be obtained from the scaled physical density by a transformation

$$\zeta = 2|v| \, (t(z-1))^{\frac{1}{z-1}}. \tag{3.38}$$

The result (the Thaler-Dynkin limit theorem) reads

$$\rho(\zeta) = \frac{z-1}{\pi}\sin\left(\frac{\pi}{z-1}\right)\frac{1}{1+\zeta^{z-1}}. \tag{3.39}$$

A second stable limit distribution can be obtained by transforming $\rho(|v|)$ itself. The infinite invariant density $\bar{\rho}(|v|)$ is related to the physical probability distribution $\rho(|v|)$ via

$$\rho(|v(t)|) \sim t^{\frac{2-z}{z-1}}\bar{\rho}(|v|). \tag{3.40}$$

It gives in the limit of $t \to \infty$

$$\bar{\rho}(|v|) \sim |v|^{1-z}. \tag{3.41}$$

This function is not normalizable. However, it will now be shown that it yields the correct scaling for some moments of the probability density. Both of these limit theorems are necessary for understanding the scaling of $Y(t)$ and $Z(t)$.

First consider the process $Y(t)$, which determines the intensity of the Moses effect in the system. It is calculated by summing the absolute values of the map

$$\langle Y(t)\rangle = \sum_{s=1}^{t}\langle|v(s)|\rangle = \sum_{s=1}^{t}\int_0^1 \mathrm{d}|v|\ |v|\rho(|v|,s). \tag{3.42}$$

The sum just adds a '+1' to the exponent obtained from the scaling of $\langle|v(s)|\rangle$. For $z > 3$ the density $\rho(|v|,t)$ can be rescaled as shown in equation (2.86). This generates a term $t^{-1/(z-1)}$, which can be shifted outside the integral

$$\langle Y(t)\rangle \sim t\int \mathrm{d}\zeta\ \left(\zeta t^{-\frac{1}{z-1}}\right)\rho(\zeta) = t^{\frac{z-2}{z-1}}\int \mathrm{d}\zeta\ \zeta\rho(\zeta). \tag{3.43}$$

The scaling of $Y(t)$ is thus given by the factor $t^{-1/(z-1)}$ multiplied by t due to the sum. For $2 < z < 3$, the first moment of expression (3.39) diverges. Instead, the scaling can be calculated using the infinite invariant density (3.41). This is done by plugging equation (3.40), which describes the scaling of the infinite invariant density, into equation (3.42). This yields

$$\langle Y(t)\rangle \sim t\int \mathrm{d}|v|\ |v|\left(t^{-\frac{2-z}{z-1}}\bar{\rho}(|v|)\right) = t^{\frac{1}{z-1}}\int \mathrm{d}|v|\ |v|\bar{\rho}(|v|), \tag{3.44}$$

which again gives an explicit time dependence. For $z < 2$ the scaling is trivial because $\rho(|v|)$ is stationary and integrable. From these considerations the scaling of $Y(t)$ is given as:

$$\langle Y(t)\rangle \propto \begin{cases} t & \text{if } z < 2 \\ t^{\frac{1}{z-1}} & \text{if } 2 < z < 3 \\ t^{\frac{z-2}{z-1}} & \text{if } 3 < z \end{cases}. \tag{3.45}$$

According to equation (3.21), the Moses effect can be calculated by subtracting $1/2$ from these results in all regimes. The resulting Moses exponents are

$$M = \begin{cases} 0.5 & \text{if } z < 2 \\ (1.5 - 0.5z)/(z-1) & \text{if } 2 < z < 3 \\ (0.5z - 1.5)/(z-1) & \text{if } 3 < z \end{cases}, \tag{3.46}$$
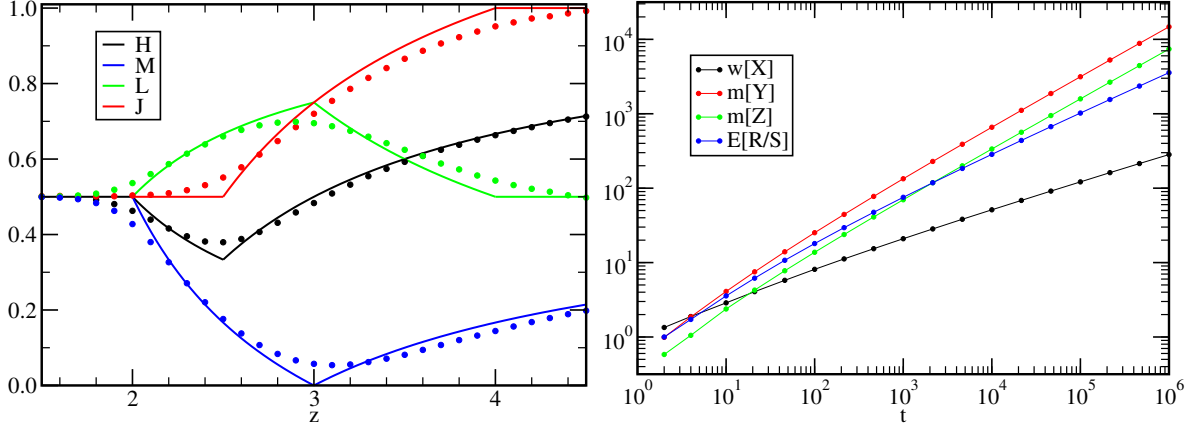
Figure 3.3

LEFT: Values of the four scaling exponents for the Pomeau-Manneville map as a function of parameter z. The circles represent the numerical results, while the theoretical predictions given in the equations (3.37), (3.46), (3.49), and (3.50) are drawn as lines. As $z \to \infty$, $H \to 1$ and $M \to \frac{1}{2}$. RIGHT: Numerical calculation of all exponents for $z = 2.5$. In the log-log plot of the width of X, median of Y and of Z, and mean of R/S are drawn as a function of time t. The scaling of these functions, that defines the four exponents can be fitted from this data. Lines connecting the data points are shown as guides to the eye.

The same methods can be used in order to find the scaling behavior of $Z(t)$, which was calculated in [C, E]. The ensemble average of $Z(t)$ is given as

$$\langle Z(t) \rangle = \sum_{s=1}^{t} \langle v^2(s) \rangle = \sum_{s=1}^{t} \int_0^1 \mathrm{d}|v| \; v^2 \rho(|v|, s). \tag{3.47}$$

This expression is integrable with respect to the infinite invariant density for $z < 4$ and with respect to the Thaler-Dynkin limit theorem for $z > 4$. Rescaling the value $v(t)$ for $z > 4$ again yields $\zeta = 2|v| \, (t(z-1))^{\frac{1}{z-1}}$. However, the time dependence of the second moment $\langle v^2(s) \rangle$ is squared compared to $\langle |v(s)| \rangle$. In contrast to this, the infinite invariant density can be plugged in just as it was done in equation (3.42). The result is

$$\langle Z(t) \rangle \propto \begin{cases} t & \text{if} \quad z < 2 \\ t^{\frac{1}{z-1}} & \text{if} \quad 2 < z < 4 \\ t^{\frac{z-3}{z-1}} & \text{if} \quad 4 < z \end{cases}. \tag{3.48}$$

This yields the following behavior of the latent exponent

$$L = \begin{cases} 0.5 & \text{if} \quad z < 2 \\ (z - 1.5)/(z-1) & \text{if} \quad 2 < z < 3 \\ 1.5/(z-1) & \text{if} \quad 3 < z < 4 \\ 0.5 & \text{if} \quad 4 < z \end{cases}, \tag{3.49}$$

when using equation (3.22), i.e. $Z(t) \sim t^{2L+2M-1}$, and the results for the Moses effect given in equation (3.46). Since now three of the four exponents are determined in

equation (3.37), equation (3.46), and equation (3.49), the Joseph exponent J can be calculated from the scaling relation $J = H - L - M + 1$. This yields

$$J = \begin{cases} 0.5 & \text{if} \quad z < 2.5 \\ (1.5z - 3)/(z - 1) & \text{if} \quad 2.5 < z < 4 \\ 1 & \text{if} \quad 4 < z \end{cases} . \tag{3.50}$$

The results were verified numerically by Kevin Bassler in [E]. The plots from the paper are shown in figure 3.3. Here for each value of z from 1.5 to 4.5 in steps of 0.1, the calculations were performed with an ensemble of 10^5 realizations of the process $x(t)$. The length of the trajectories was $t = 10^6$ map iterations. The exponent H was estimated by the scaling of the difference between the third quantile and the first quantile of the ensemble for each time step. For the scaling, a power law was fitted to the results for different times. When fitted in a log-log plot these power laws are straight lines. The Joseph exponent J is estimated using the method of R/S statistics. This is known to be equivalent to DFA0 [65]. The two series $Y(t)$ and $Z(t)$ that determine the Moses exponent M and the Latent exponent L, are calculated as defined in the equations (3.20) and (3.19). However, instead of the ensemble average, the median is taken because results seem to be more stable here.

As an example, the calculation for $z = 2.5$ is shown in the right panel of figure 3.3. The statistical error of the data points is smaller than the symbol size, however, the systematic deviation of the scaling for short times from the asymptotic behavior leads to some error compared to the theoretical results. These are especially strong close to the crossover points. The curvature is clearly visible in the figure. Results can be improved by taking longer trajectories into account. In the left panel of the figure, the numerical results for all considered values of z are shown. The simulation results roughly follow the theoretical predictions with the already mentioned deviations around the crossover points.

For parameter values $z < 2$ the system exhibits normal diffusion. This is despite already in this regime some interesting effects happen that are studied e.g. in [85]. However, the density of the symmetric map is stationary and has a well-defined mean. For $z > 2$ the density becomes non-stationary as described in section 2.4.4. It slowly approaches $v = 0$. Here, due to the fat tails of the rescaled probability distribution (3.39), the mean goes to zero more slowly than the density itself. This leads to a Noah effect in addition to the Moses effect, which is caused by the aging behavior of the density. For $z > 3$ this is still the case, however, the distribution shown in (2.83) is steeper here. Therefore, the mean goes to zero with the same scaling as the distribution itself, causing a decrease of the Moses effect.

Surprisingly, a Joseph effect appears even though the process is renewed after each waiting period. This is due to the diverging mean waiting time. However, the Joseph effect is only present for $z > 2.5$, which implies that for $2 < z < 2.5$ the correlation time is finite despite the divergence of the mean waiting time. The stronger the effect of aging and the longer the maximal waiting times get for larger values of z, the stronger also the Joseph effect becomes. It gets its maximal value of $J = 1$ already for $z = 4$. Both the Noah and the Joseph effect contribute to an increase in H, while the Moses

effect in the current setting leads to subdiffusion. For $z > 3$ the combination of Joseph and Noah effect overpowers the Moses effect.

3.3 Further observations

For some classes of processes, general statements can be made about the decomposition of the Hurst exponent H into the three exponents J, L and M. Common settings lead to specific combinations of several effects. In section 3.3.1 examples for such settings are discussed. One additional example of a special class contains jump models like continuous time random walks [112]. They are discussed in section 3.3.3. A second class describes one-sided (asymmetric) processes. Here the effects are exemplified by the standard Pomeau-Manneville map in section 3.3.4.

3.3.1 Interpretation of the Moses and the Noah effect

The Noah effect and the Moses effect are both properties of the probability distribution $\rho(v, t)$. The paradigmatic models in section 3.1.3 explain the two most simple special cases of scaling. The Pomeau-Manneville map illustrates how infinite densities manifest themselves in the framework of Noah and Moses effects. They lead to several configurations where one or both of the exponents differ from ¹/₂.

The most simple case is explicit time dependence of the increments. Here, in order to get meaningful results, one simply needs to rescale the increments. This always manifests itself in a Moses effect, because

$$\langle |v(t)| \rangle = t^\gamma \langle |\zeta(t)| \rangle \;\Rightarrow\; M = \gamma + \frac{1}{2}, \tag{3.51}$$

where $\langle |\zeta(t)| \rangle$ is stationary and integrable. There is no Noah effect in this case, because

$$\langle v^2(t) \rangle = t^{2\gamma} \langle \zeta^2(t) \rangle \;\Rightarrow\; L = \frac{2\gamma}{2} - M + 1 = \frac{1}{2} \tag{3.52}$$

is consistent with $L = $ ¹/₂.

What if the time dependence of the increment distribution is not explicit but due to waiting times with diverging mean? In this case, there exists a regime for both the first and the second moment, where the infinite invariant density $\bar{\rho}(v) = t^\gamma$ has to be used for calculations. For the Moses effect this again leads to

$$\langle |v(t)| \rangle = t^\gamma \int \mathrm{d}v \; |v| \bar{\rho}(v) \;\Rightarrow\; M = \gamma + 1/2. \tag{3.53}$$

The divergence of the mean waiting time leads to aging. The probability of the system being in a long waiting period increases over time. Therefore, a Moses effect is expected here. However, this case is different from the scaled increment distribution discussed above. The difference is visible in the exponent L. It can be calculated as

$$\langle v^2(t) \rangle = t^\gamma \int \mathrm{d}v \; v^2 \bar{\rho}(v) \;\Rightarrow\; L = \frac{\gamma}{2} - M + 1 = \frac{3}{4} - \frac{M}{2}. \tag{3.54}$$
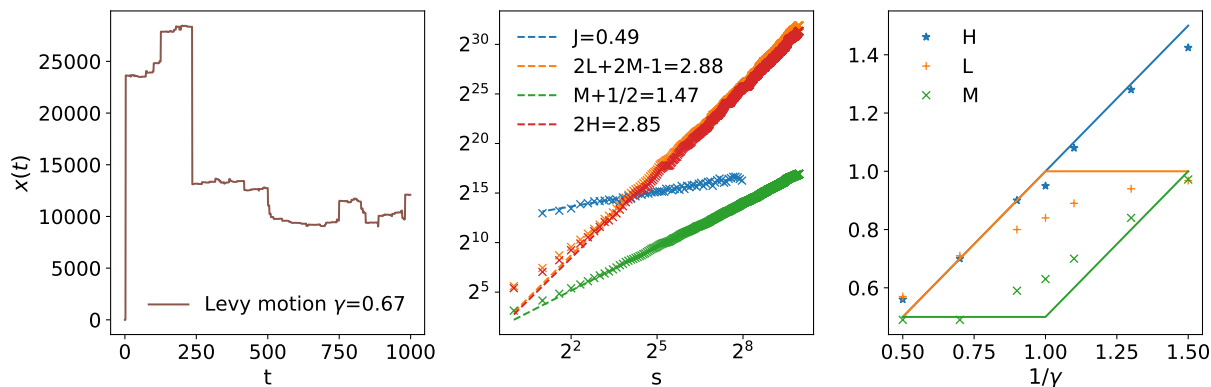
Figure 3.4

Lévy motion with extremely fat tails. LEFT: Trajectory of a process with symmetrically distributed increments with tails that decay with $|v|^{1/\gamma}$ and $\gamma = 2/3$. CENTER: Numerical calculation of the four exponents in such a system, using 1000 trajectories and the median instead of the ensemble average. The System exhibits $H = 1.43$, which is close to the expected value of 1.5. It is caused by both a Noah effect and a Moses effect. RIGHT: Results of numerical calculations for H (blue), L (orange), and M (green) displayed as crosses. The results roughly follow the predictions indicated by the lines of the same color.

So there is always a Noah effect present in the opposite direction to the Moses effect with half its deviation from $1/2$. This can be interpreted as follows. While the physical probability distribution $\rho(v, t)$ does not have fat tails, the scaled infinite invariant density has. In this way the time-dependent factor, which enters due to the Moses effect, generates a Noah effect. The relation between the two exponents is always fixed, therefore measuring the characteristic relation, $L = 3/4 - M/2$, is an indicator for an infinite invariant density.

A different example is illustrated by the Pomeau-Manneville map in the regime $3 < z < 4$. For this range of parameters, the first moment and the second moment are integrable with different limit theorems. In this case, the infinite invariant density can be used for one calculation while the other calculation is done with a rescaled variable. This leads to different combinations of the Moses effect and the Noah effect. However, the interpretation is the same as before. The Noah effect is a consequence of the Moses effect and not a direct property of the physical density.

One more special case should be discussed in detail. Consider the paradigmatic model for the Noah effect presented in section 3.1.3. Here the increments v are fat tailed, so the second moment of the density diverges. The scaling is given by a Lévy central limit theorem as

$$\rho(x(t)) = t^H \rho\left(x(t)/t^H\right). \tag{3.55}$$

Here the scaling of the process Z can not be calculated by rescaling v, because $\rho(v)$ is constant. The scaling is a consequence of the diverging second moment. The first moment, however, is integrable with respect to the physical density $\rho(v)$. So $Y(t)$ scales

linearly and the Moses exponent is $M = \frac{1}{2}$. As it was described in section 2.4.3, the scaling of a distribution $\rho(x)$ with fat tails only depends on the power of x in the tails. So once the scaling of $\rho(x) = \rho(\sum_s v(s))$ is known, the scaling of $\rho(\sum_s v^2(s))$ is straightforward. Since the values v are uncorrelated,

$$\langle v(t)v(t + \Delta)\rangle = \delta(\Delta), \tag{3.56}$$

and $\rho(x(t)/t^H)$ is a stable distribution, it follows

$$\rho\left(\sum_{s=1}^{t} v^2(s)\right) \sim \rho\left(\left(\sum_{s=1}^{t} v(s)\right)^2\right) = \rho\left(x^2(t)\right) \Rightarrow \rho\left(\frac{\sum_{s=1}^{t} v^2(s)}{t^{2H}}\right) \text{ stable .} \tag{3.57}$$

So, using the rescaled variable $\zeta = (\sum_s v^2(s))/t^{2H}$, the latent exponent L can be calculated from

$$\langle Z(t)\rangle = \left\langle \sum_{s=1}^{t} v^2(s)\right\rangle = t^{2H}\langle\zeta\rangle = t^{2H}\int \mathrm{d}\zeta\, \zeta^2 \rho(\zeta) \propto t^{2H}. \tag{3.58}$$

This calculation contains one major problem. The mean $\langle\zeta\rangle$ is not defined due to the fat tails of the distribution. Instead of using the ensemble average $\langle Z\rangle$, it turns out to be beneficial to use the median of $Z(t)$ for the calculation. This leads to nice scaling in the examples shown in figure 3.1 and figure 3.4. With $M = \frac{1}{2}$, from equation (3.58) follows

$$L = H - M + \frac{1}{2} = H. \tag{3.59}$$

So the system exhibits the expected behavior. The Hurst exponent can be explained solely by a Noah effect.

Now what if the tails of the distribution are even heavier, such that the first moment already diverges? In this case $Y(t)$ has to be calculated in the same way as $Z(t)$ using the Lévy central limit theorem. The tail of the distribution of the sum of the absolute values $|v|$ has the same decay γ as the distribution of the values v themselves. Therefore, the rescaled variable with a stable probability density is

$$\zeta' = \frac{\sum_{s=1}^{t} |v(s)|}{t^H}. \tag{3.60}$$

Accordingly the calculation reads

$$\langle Y(t)\rangle = \left\langle \sum_{s=1}^{t} |v(s)|\right\rangle = t^H\langle\zeta'\rangle = t^H\int \mathrm{d}\zeta'\, \zeta' \rho(\zeta') \propto t^H. \tag{3.61}$$

Again, like equation (3.58), this calculation is not rigorous and using the median is recommended instead of the ensemble average. In this case, where the first moment of the increment distribution diverges, the Moses effect is

$$M = H - \frac{1}{2}. \tag{3.62}$$

This directly follows from equation (3.61). Since $H > 1$ always in this scenario, the Moses effect is always greater than $1/2$ and contributes to larger Hurst effects, i.e. superballistic diffusion. The process $Z(t)$ behaves in the same way as above where only the second moment diverges. It scales with $Z(t) \propto t^{2H}$. Therefore equation (3.59) can be modified with the result for the Moses effect. This yields

$$L = H - M + 1/2 = H - (H - 1/2) + 1/2 = 1, \tag{3.63}$$

a constant Noah effect with $L = 1$. This is the maximum value that L can obtain (similar to J). Numerical results for this case are presented in figure 3.4. In the left panel, a trajectory of a process with diverging first moment is shown. In the central panel, numerical calculations of the exponents are presented. The right panel shows the results of several numerical calculations of systems with diverging or finite first moment are shown. They roughly follow the theory presented above.

It might sound counter-intuitive that the paradigmatic model for the Noah effect can also generate a Moses effect for extreme values. However, the Moses effect is designed to detect aging in $x(t)$. Lévy motion with diverging first moment of the increments indeed exhibits aging. The reason is similar to the effect of aging in the Pomeau-Manneville map. Since the increments $v(t)$ can be extremely large the total value of $x(t) = \sum_{s=1}^{t} v(s)$ is dominated by the single biggest jump. The longer the system runs the more likely such an extreme jump is to occur.

The latent exponent L and the Joseph exponent J can both not exceed 1. This is because extremely fat tails and long memory lead to aging eventually. Therefore, a Moses effect arises.
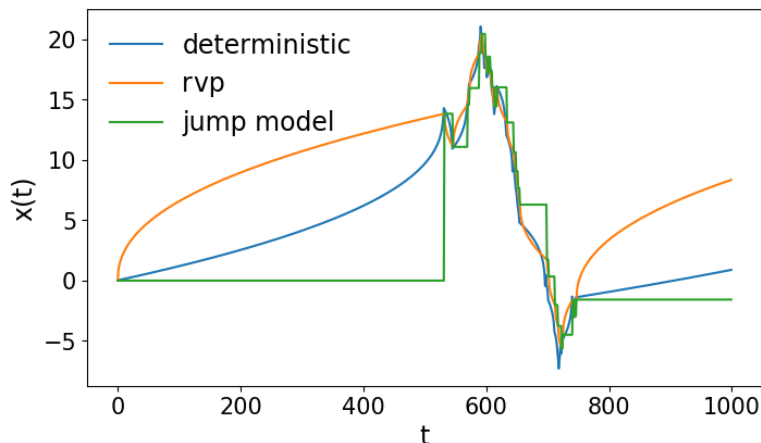
3.3.2 Renewal velocity processes

The relation between the exponent of the waiting times γ and the corresponding jump lengths coupled by the exponent β in a stochastic renewal process and the parameter z of the Pomeau-Manneville map was derived in section 3.2.2. The dynamics of the map is approximated for $\gamma = 1 - \beta$. The stochastic models are subject of work in progress [J]. Here only three remarks about these systems shall be included.

Firstly, the stochastic models allow the exponent of the waiting time distribution γ and the exponent β to be varied independently. The result is a two-dimensional phase space that is shown in [7] and [4]. It can be regarded as a generalization of the systems with only the Pomeau-Manneville parameter z that are discussed here. The Joseph effect can even be found for finite mean waiting times $\gamma > 1$ if the parameter β is sufficiently large [7].

Secondly, the stochastic models are continuous in time. Therefore the definition of what an increment is has to be adjusted. Since the theory discusses only the long time scaling of the functions $\langle x^2 \rangle$, $\langle \sum_t Y^2(t) \rangle$, $\langle \sum_t Z^2(t) \rangle$, and $\langle \overline{\delta^2} \rangle$, the results do not depend on the exact way how to make values discrete.

Thirdly, the shape of the path during the waiting times can be regarded as an additional parameter [20]. A stochastic process with waiting times and a fixed velocity,

Figure 3.5

Trajectories for the jump model (3.65), the 'renewal velocity process' (rvp) defined in equation (3.64), and deterministic diffusion generated by the Pomeau-Manneville map (3.29). After each waiting period in which the exact paths differ the systems have identical values for the randomly generated waiting times are the same. This is true exactly for both stochastic systems. It is approximately true for the trajectory generated by the Pomeau-Manneville map, since here the short chaotic dynamics can lead to small deviations. The figure was published in [C].

which only switches its sign, is called a Lévy walk. A stochastic process with velocities, which are constant during the waiting time, but depend on its duration, is a generalized Lévy walk [7]. In [C] this parameter was chosen in such a way that the velocity during the waiting time is described by

$$v(t) \propto \gamma(t - t_n)^{\gamma - 1}. \tag{3.64}$$

A trajectory is shown in figure 3.5. A generalized Lévy walk or a process of the type described in equation (3.64) can model the dynamics of the Pomeau-Manneville map if the walk lengths are coupled to the waiting times as described in equation (3.36). Here the results for all scaling exponents are the same as for the map. They are independent of the exact paths. The following section will deal with a similar class of systems where $x(t)$ (the particle) rest during the waiting time and performs a jump at its end.

3.3.3 Jump models

In figure 3.5 three processes are shown. All of them are displaced equally far during the waiting times and the probability distribution of their waiting times is identical. One is the Pomeau-Manneville map and one is a special case for a renewal velocity process, which was introduced above. The third example is a jump process. The three processes were compared in [C]. As mentioned in section 3.2.2, the Hurst exponent is identical for all three processes, since the position after the jump is also identical.

The exact path of the renewal velocity process does not influence the four exponents H, L, M and J. However, the extreme case of jump models is indeed different. Here x is
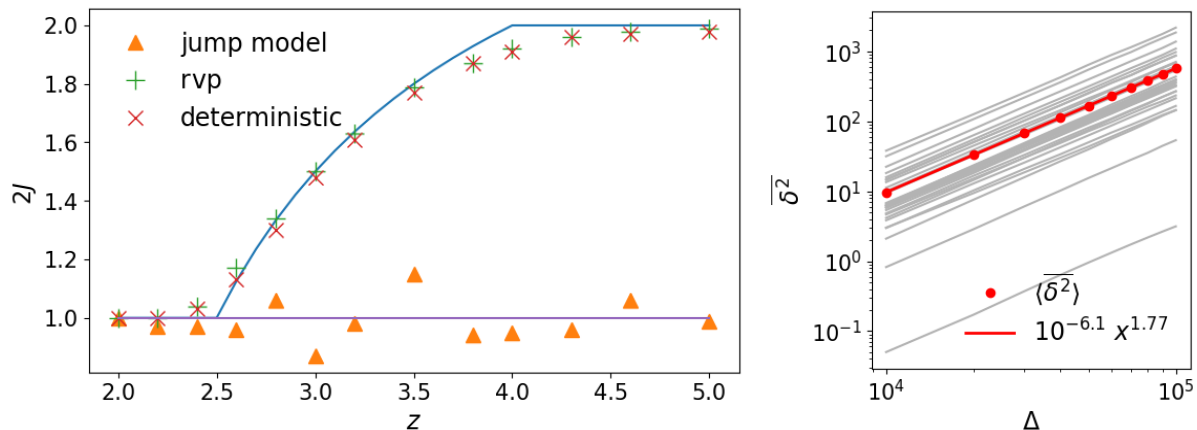
Figure 3.6

LEFT: Scaling exponents $2J$ of the TA MSD for the jump model (3.65) the renewal velocity process (3.64) and deterministic diffusion generated by the Pomeau-Manneville map (3.29). The jump process always yields $J = 1/2$, while the other model both follow the theory described in equation (3.50). RIGHT: Scaling of the TA MSD of the deterministic system driven by the Pomeau-Manneville map for different realizations. Despite the time average is a described by random values, its scaling with the lag time Δ is the same for different realizations. Both parts of the figure were published in [C].

displaced instantaneously after the waiting period and rests during the waiting period. For jump models with non-continuous paths, the exponent J is always $1/2$. The following derivation was performed by Philipp Meyer and published in [C].

The velocity in a jump model can be written in the (not rigorous) form

$$v(t) = \sum_{\{i:t_i<t\}} \chi_i \delta(t - t_i), \tag{3.65}$$

with the well known δ-distribution. The asymptotic scaling of the velocity displacement defines M and L, so $\langle v^2(t) \rangle \propto t^{2M+2L-2} = t^\beta$.

It can be shown that the TA MSD of such a jump process always scales linearly with Δ. This is because the EA MSD,

$$\langle x^2(t) \rangle = \left\langle \int_0^t dt_1 v(t_1) \int_0^t dt_2 v(t_2) \right\rangle, \tag{3.66}$$

can be rewritten with the θ-step function as the integral over $\delta(t)$

$$\langle x^2(t) \rangle = \left\langle \sum_{ij} \chi_i \chi_j \theta(t - t_i) \theta(t - t_j) \right\rangle. \tag{3.67}$$

Then using $\langle \chi_i \chi_j \rangle = \delta_{ij}$, this yields

$$\langle x^2(t) \rangle = \left\langle \left(\int dt v(t) \right)^2 \right\rangle = \int dt \langle (v(t))^2 \rangle \propto t^{\beta+1}, \tag{3.68}$$

and therefore $2H = \beta + 1$. This implies $J = 1/2$. These results were derived previously in [3] for stored energy driven Lévy flights (which is exactly the model with jumps introduced here). It was calculated from the theory of continuous time random walks.

The results for the scaling of the EA TA MSD in three systems are numerically validated in figure 3.6 in the left panel. They are compared for parameters z or the corresponding values of γ and β. While the process generated by the Pomeau-Manneville map and the renewal velocity process scale with the same exponent J, the jump model does not. Here the Joseph exponent is always $1/2$.

Despite the derivation presented in section 3.1.1 does not necessarily hold for jump processes, the scaling relation still seems to be valid for symmetric jump processes [63]. Since $J = 1/2$ it simplifies to

$$H = M + L - 1/2. \tag{3.69}$$

A natural question that remains is, how typical this expected behavior of the TA MSD really is. It is not always possible to create an ensemble and therefore the framework would be more meaningful if the scaling could already be obtained from one realization of the process. It is here only tested numerically for the process generated by the Pomeau-Manneville map. The result is shown in the right panel of figure 3.6. Indeed all tested realizations yield almost identical scaling of the TA MSD with lag time. The difference is only a pre-factor and is not in the scaling. So the Joseph effect, at least in this case, can be obtained correctly without an ensemble. The same is not true for the scaling with measurement time t, at least for the Pomeau-Manneville map. This was demonstrated numerically in [C].

3.3.4 Asymmetric processes

This final section of the chapter about the Joseph, Noah, and Moses effects shall give one perspective to a possible extension of the theory. So far only symmetric processes were considered. This makes sense from the perspective of free diffusion. Here at least one additional effect can be seen by looking at one-sided distributions of $\rho(v)$. Many real processes, in fact, can not be negative such as precipitation or volcanic activity.

An interesting property of the definitions of the Noah and the Moses effect is that they do not take into account the sign of v. The functions Y and Z are defined for the absolute value or the square of v. Accordingly, both exponents L and M have to be unchanged if a process $|v(t)|$ is considered instead of $v(t)$.

It is easy to see from the definition of the TA MSD (3.10),

$$\overline{\delta^2(\Delta)} = \overline{(x(t+\Delta) - x(t))^2} = \frac{1}{T-\Delta} \sum_{t=1}^{T-\Delta} (x(t+\Delta) - x(t))^2, \tag{3.70}$$
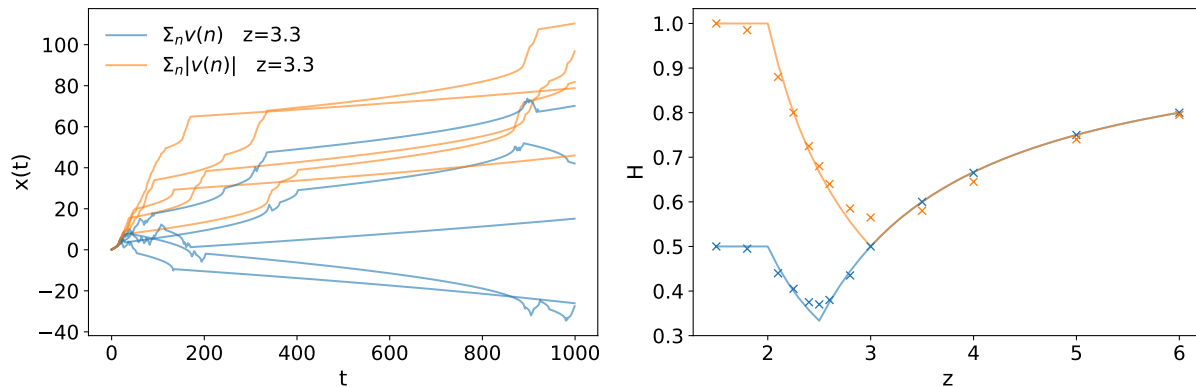
Figure 3.7

The symmetric version of the Pomeau-Manneville map versus the asymmetric version. LEFT: five trajectories with 1000 iterations for diffusion generated by the symmetric (blue) and the asymmetric map (orange). The spreading of the ensemble indicated the mean squared displacement. RIGHT: Hurst exponents H of the EA MSD for the symmetric (blue) and the asymmetric (orange) map. The exponent is calculated for several values of z and compared with the respective theory curve (solid lines) of equation (3.37) (blue) and equation (3.71) (orange). The values seem to match and only deviate a bit more around the crossover points.

that even for independent stationary increments the ensemble average of the expression in the integral would scale like $\mu^2\Delta^2$ with the mean value μ of the increments. So the Joseph exponent would already be $J = 1$ in this case. Since 1 is also the maximal value that J can take this result implies that $J = 1$ is valid everywhere for one-sided processes.

Now this knowledge can be used for understanding the difference between positive and symmetric increments for more complicated systems like the Pomeau-Manneville map. The positive version was introduced in section 2.4.4 while the symmetric version was discussed in section 3.2. For the symmetric version all exponents were calculated. They can be found in the equations (3.37), (3.46), (3.49) and (3.50). As mentioned above, the calculation of L and M is equivalent for the asymmetric version of the map. The Joseph effect is expected to be $J = 1$ for the previously discussed reasons. Under the assumption that the scaling law (3.23) still holds, H could be calculated from these numbers. The result is

$$H = \begin{cases} 1 & \text{if} \quad z < 2 \\ 1/(z-1) & \text{if} \quad 2 < z < 3 \\ (z-2)(z-1) & \text{if} \quad 3 < z \end{cases} . \tag{3.71}$$

It is ballistic only for $z < 2$ and then becomes superdiffusive for $z > 2$ due to the Moses effect. For $z > 3$ the Hurst exponent grows again because the Moses exponent does, too. This is analogous to the symmetric map. For $z > 3$ the result is identical to the symmetric case. This sounds counter-intuitive due to the fact that one system

diffuses while the other one always grows. However, the analytical results derived in [93] confirm equation (3.71). A numerical check is shown in figure 3.7. The equivalence of the scaling of both systems for large z is due to the dominance of a single long waiting period associated with the biggest jump in one direction. The sign of the smaller jumps during shorter waiting periods does not contribute to the scaling.

The EA MSD of a monotonous process generated by positive increments can be understood by the scaling relation (3.23) at least in the case of the Pomeau-Manneville map. The modified scaling relation for asymmetric processes with $J = 1$ reads

$$H = M + L. \tag{3.72}$$

# 4 Short range correlations and detrended fluctuation analysis

Out of the three effects discussed in the previous section, one was repeatedly observed in atmospheric data sets. The Joseph effect was reported for global mean surface temperatures [107], local temperature measurements [39], and precipitation data [64]. A popular method for the analysis is DFA.

There is a long history of investigations on performance and pitfalls of this method [58, 132, 130, 82, 28, 23, 134]. Some solutions for special cases were calculated [121, 53, 13, 29, 54]. The interpretation of the observed long range correlations is not easy because, for finite time series, long range correlations and short range correlations are difficult to distinguish. Accordingly, scientists have reported that, for example, DFA results for temperature are compatible with the autoregressive model of order one (AR(1)) with exponentially decaying correlations [82].

Recently, Höll and Kantz [55] developed a theoretical description of DFA by establishing a direct relation to the correlation function, which enables a better understanding of its results (see section 2.3.3). Here these results are used to analyze real data with more care than in traditional usage of DFA. In addition, some further insights into the DFA fluctuation function are developed. Section 4.1 summarizes what is known or expected from fluctuation functions. Section 4.2 includes a case study for global mean surface temperatures and shows that modeling these temperatures with long range correlated models is unnecessary. Section 4.3.1 goes one step further and discusses the possibility of fitting measured DFA fluctuation functions with theoretical fluctuation functions.

4.1 Theoretical fluctuation functions

In order to detect long range correlations, many scientists have employed DFA. The only property of interest for their studies is the asymptotic scaling of the fluctuation function. In contrast to this usage, this section shall provide an overview of what DFA really measures. Such a discussion is easier since the relation between the fluctuation function and the correlation function was found [55]. Several properties are discussed like the exact meaning of the exponent α compared to the Joseph exponent J. Moreover, the exact fluctuation function of the most simple short range correlated process and the result of combinations of simple processes are investigated. A similar but far more compact discussion was published in [F].
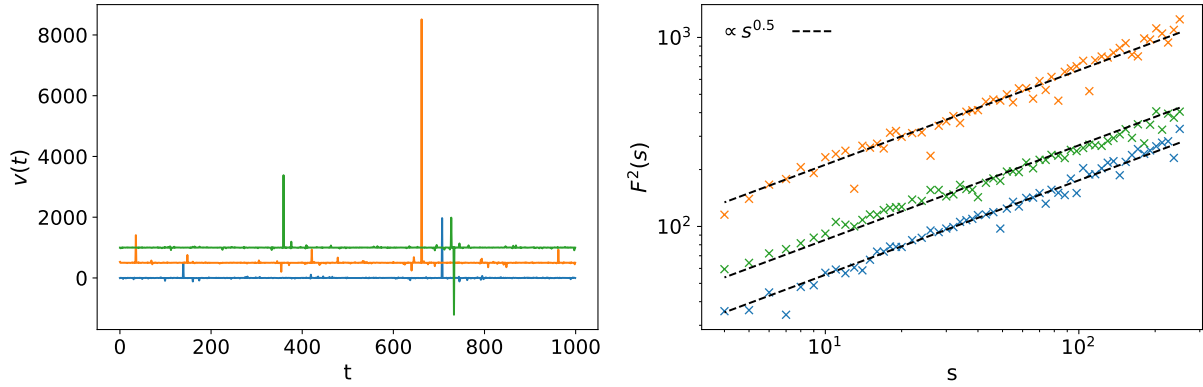
Figure 4.1

LEFT: Three trajectories of an i.i.d. process with power law increments with random signs. The asymptotic scaling of the density is v^{-1}. RIGHT: DFA for the tree time series. The scaling is $\alpha = 1/2$ indicated by the dashed lines. The pre-factor depends on the measured variance and is therefore different for the three realizations.

4.1.1 Meaning of the DFA exponent

In the previous chapter, it was shown that the Hurst exponent H, which is found for anomalous diffusion, can be decomposed into three root causes, namely the Joseph, Noah, and Moses effect. It was already discussed in chapter 2 that the scaling exponent of DFA0 is equivalent to the exponent of the TA MSD. Since the ensemble mean of the TA MSD is exactly the property that defines the Joseph exponent, DFA0 is an alternative measure for J

$$\alpha = J \quad \text{for DFA0.} \tag{4.1}$$

The Joseph exponent measures the autocorrelations of the time series, as does the exponent α in DFA. For stationary time series it was shown that the fluctuation function for an arbitrary order q can be expressed as a transformation of the autocorrelation function. The relation reads, according to equation (2.59),

$$\langle F_v{}^2(s) \rangle = \sigma^2 \left(L(0, s) + 2 \sum_{t=1}^{s-1} C_{vv}(t) L(t, s) \right). \tag{4.2}$$

The equation above is still valid if the probability distribution of the values $v(t)$ in the generating process has a diverging second moment, i.e. the Noah effect is present. This is easy to understand since the second moment of a measured time series can always be calculated and is always finite. If there is a Noah effect in the process this means that the variance σ^2 can not be predicted analytically and in fact becomes a random variable that depends on the realization of the process.

An example of the DFA result of a model that exhibits the Noah effect is shown in figure 4.1. A similar example was published in [F]. It was not yet investigated how a
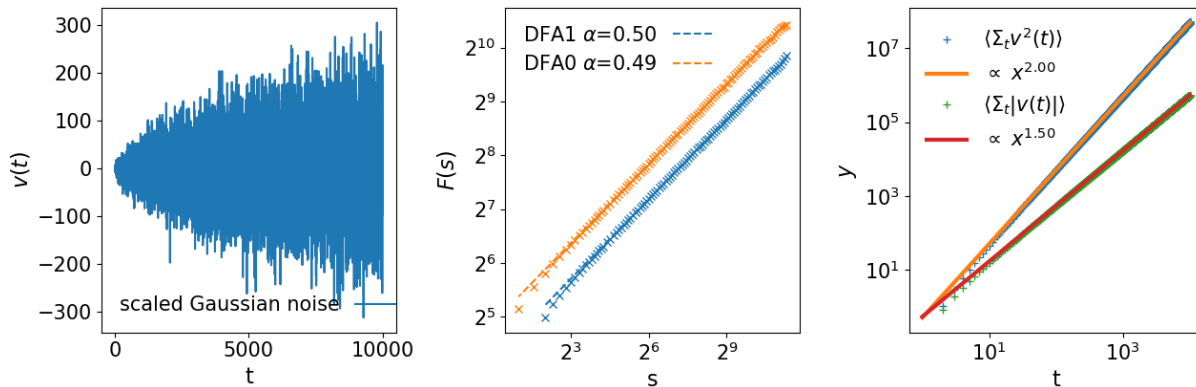
Figure 4.2

LEFT: Trajectory of scaled Gaussian noise generated as Gaussian random variables with a time dependent pre-factor $t^{1/2}$. CENTER: DFA0 and DFA1 for processes of the same type. Each value of the fluctuation function was obtained as the mean of the analysis for 100 realizations. Both DFA0 and DFA1 yield $\alpha = {}^1\!/_2$. RIGHT: numerical calculation of the Noah effect and the Moses effect with an ensemble of 1000 realizations. The scaling of $\langle \sum_t |v(t)| \rangle$ yields $M = 1$. By comparison with the scaling of $\langle \sum_t v^2(t) \rangle$ L is determined to be $L = 0.5$.

Noah effect affects the performance of DFA in terms of expected uncertainties in the measured exponents. Presumably, it has a negative effect, since very few extreme events dominate the result.

So for stationary time series, DFA is a measure of the Joseph exponent only. For data with an external trend, DFAq with $q > 0$ deviates from DFA0. However, the exponents H, J, L and M are difficult to define here, so the case shall not be discussed in this context. More interesting is the question: what happens if the dynamics is non-stationary due to a Moses effect? Here two cases should be distinguished. They can be exemplified by scaled Gaussian noise and Brownian motion representing the two cases of non-stationary fluctuating dynamics and diffusive dynamics.

Scaled Gaussian noise is a process that consists of independent Gaussian random variables $v(t)$. In contrast to Gaussian white noise, the variance of the probability distribution is time-dependent. More specifically, it exhibits a power law time-dependence. The integral over these random variables would then yield scaled Brownian motion which is a popular model of anomalous diffusion [63]. Discrete scaled Gaussian noise can be generated by a simple multiplication of the scaling factor to white noise $\epsilon(t)$

$$v(t) = t^{H-0.5}\epsilon(t). \tag{4.3}$$

As it was already discussed in section 3.1.3, this process is the paradigmatic model for the Moses effect, where no other effects are present and therefore $H = M$. DFA for such processes indicates $\alpha = {}^1\!/_2$. This is because the values are uncorrelated and the non-stationarity only leads to an unequal distribution for the K segments in the DFA
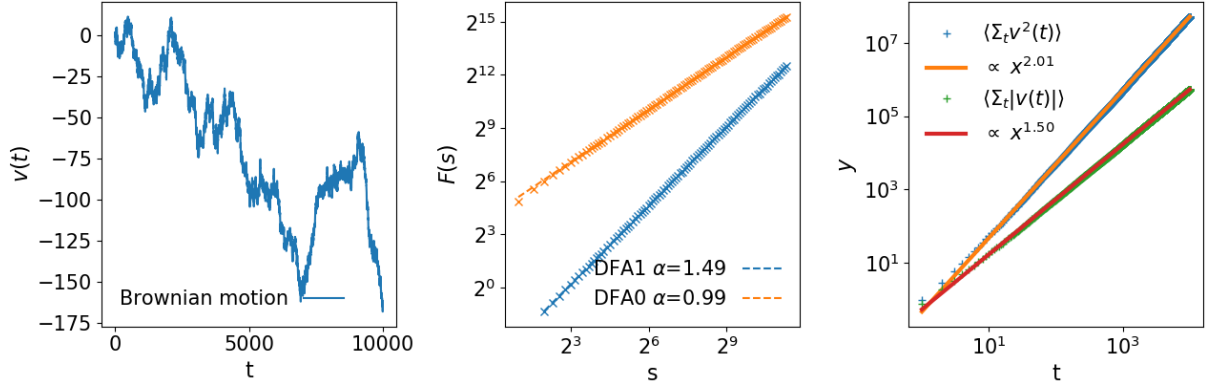
Figure 4.3

LEFT: Trajectory of Brownian motion generated as sum over Gaussian random variables. CENTER: DFA0 and DFA1 for processes of the same type. Each value of the fluctuation function was obtained as the mean of the analysis for 100 realizations. While DFA0 yields $\alpha = 0.99 \approx 1$, DFA1 yields $\alpha = 1.48 \approx 3/2$. RIGHT: numerical calculation of the Noah effect and the Moses effect with an ensemble of 1000 realizations. The scaling of $\langle \sum_t |v(t)| \rangle$ yields $M = 1.01$. By comparison with the scaling of $\langle \sum_t v^2(t) \rangle$ L is determined to be $L = 0.51$.

algorithm. Since equation (2.47) only takes the mean of the segments for each s, the final fluctuation function $F(s)$ does not see the Moses effect. This is shown in figure 4.2, where the Moses effect is numerically calculated for the case $H = 1$ and the Josef exponent is shown to be equal to the DFAq exponent.

For Brownian motion, DFAq with $q > 0$ is known to yield a scaling exponent of $\alpha = 3/2$ [53]. Here it deviates from rescaled range analysis or DFA0, which both yield $J = 1$ [65]. So, in this case, the DFA exponent is greater than the Joseph exponent. In fact, it is equal to H. So where does the difference $\alpha - J > 0$ originate? The analysis yields a Moses effect of $M = 1$ for Brownian motion. The complete analysis is shown in figure 4.3. So obviously, in this case, DFAq does see the Moses effect. The results for fractional Brownian motion are similar in this regard.

In conclusion, there must be two different kinds of Moses effects. One describes the explicit scaling of the increment distributions as it is present in scaled Gaussian noise. It could be called the Moses effect of fluctuations M_{fluct}. The other one is the Moses effect that is due to diffusion as in Brownian motion and might be denoted as M_{diff}. So in total $M = M_{\text{fluct}} + M_{\text{diff}} - 1/2$. In fact for symmetric processes without trend, the DFAq exponent seems to be equal to

$$\alpha = J + M_{\text{diff}} - 1/2. \tag{4.4}$$

This might already sound complicated, however, it is still not a sufficient description of DFA even for simple signals. As discussed in section 3.3.4 the Joseph effect for one-sided processes is always $J = 1$. This is not true for DFA, where the result of one-sided
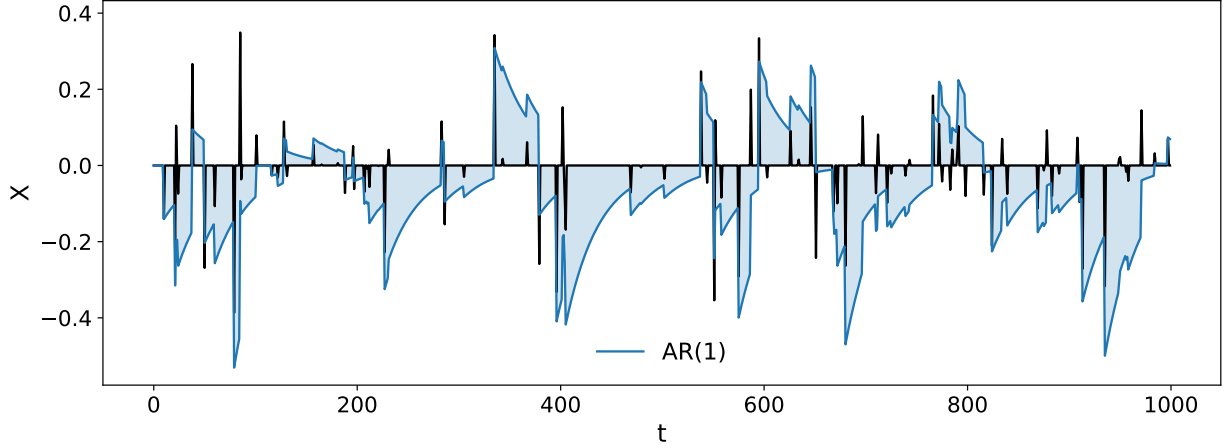
Figure 4.4
AR(1) (blue) with $a = 0.97$, driven by a realization of uncorrelated noise η (black) with $\rho(\xi) = \pm \exp(-10\xi)$ and $W(r) = \exp(-0.1r)$.

processes is always the same as for a symmetric process if all other properties are the same (see section 4.3.2).

4.1.2 Uncorrelated noise

According to the previous section, the scaling of the fluctuation function of i.i.d. variables is $\alpha = 1/2$, even if the probability distribution has infinite variance. More precisely the squared fluctuation function is given as [55]

$$F^2(s) = \sigma^2 \frac{(s^2 - 4)}{15s}. \tag{4.5}$$

So DFA does not see the distribution of the process and there is no reason to restrict any analysis to Gaussian processes. Uncorrelated noise can be generated by any discrete i.i.d. process or even more general, even if the inter-event times follow some random distribution. If inter-event times are exponentially distributed that means, that events have the same probability in every time range of a certain length. Such a (Poisson) process $\eta(t)$ is also uncorrelated in the sense

$$\langle \eta(t)\eta(t + \Delta) \rangle = 0 \quad \forall \Delta \neq 0. \tag{4.6}$$

If the observation time is not too short compared to the typical inter-event time, the process will yield the same DFA result as Gaussian white noise.

The focus in this chapter, as well as in the following chapter, is on short range correlated processes. The most simple stochastic processes that generate short range correlated data are the autoregressive processes, introduced in section 2.2.3. These models can be written as

$$X(t) = \sum_i a_i X(t - i) + \epsilon(t), \tag{4.7}$$

with the AR parameters a_i. The sum goes from one to some maximum value that is the order of the autoregressive model. In the derivation of the autocorrelation function of AR(1) and AR(2), the Gaussian shape of the distribution was not used (also see [21]). According to equation (2.59), the scaling of the DFA fluctuation function only depends on the correlation function. Therefore, the noise in the model should also be generalized if the model is compared to data with DFA. So the Gaussian noise $\epsilon(t)$ can be replaced by a general uncorrelated time series $\eta(t)$. An autoregressive process then defines an iteration function D that generates the dynamics

$$X(t) = D\Big(\eta(t), X(t-1), X(t-2), ...\Big). \tag{4.8}$$

Examples of how autoregressive dynamics on uncorrelated but non-Gaussian noise look like are shown in figure 4.4. Here a set of independent, identically distributed random variables with exponential distribution $\rho(\xi)$ is considered. These random variables are now also randomly distributed in time, where the inter-event times also have an exponential probability distribution $W(r)$.

4.1.3 Theoretical fluctuation function of AR(1)

The most simple dynamics is described by linear response models with exponentially decaying autocorrelations. In continuous time, such models are Ornstein-Uhlenbeck processes [30]. In discrete time, it is modeled by the autoregressive model of order one (AR(1), see section 2.2.3). The process is given by the iteration

$$X(t) = gX(t-1) + \eta(t). \tag{4.9}$$

The dynamics is driven by white noise η. The memory is determined by the AR-parameters g in the range $0 < g < 1$, such that the correlation function $C(t) = g^t$ decays exponentially. As discussed in section 4.1, the distribution of the noise η is not important here as long as it is uncorrelated.

It is well known in the literature that results for processes with exponential decay of correlations in DFA can easily be misinterpreted for long range correlations [82]. Thus results should be treated with care, especially when dealing with short time series. Therefore, it is helpful to know the exact shape of the AR(1) fluctuation function and compare it to the fluctuation function of the data.

When plugging the AR(1) correlation function (2.23) into equation (2.59) with a suitable kernel $L(t, s)$, one can obtain the theoretical fluctuation function of AR(1). It has the form [54]

$$F^2(s) = \sigma^2 \frac{g^s \Lambda(s) + \Xi(s)}{\varkappa(s)} \tag{4.10}$$

For DFA1 with the kernel (2.61) the result was first presented in [54]. The calculation
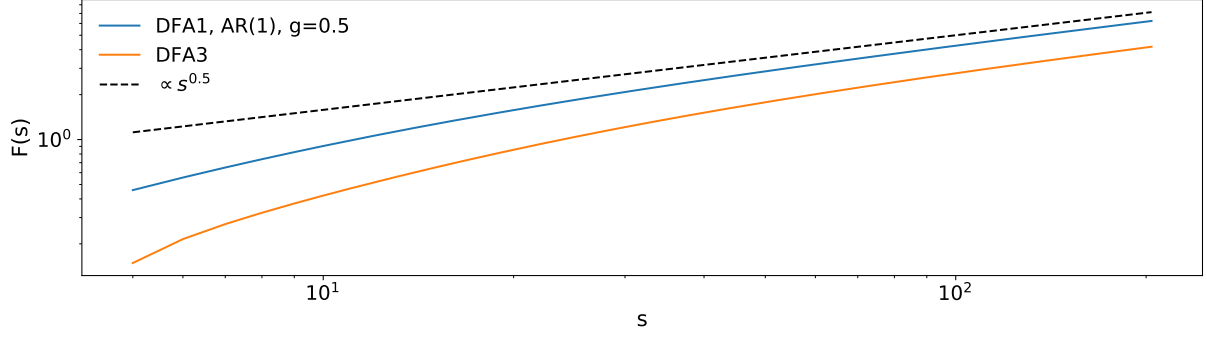
Figure 4.5
The fluctuation functions of the AR(1) process shown in equation (4.10) for $g = 0.5$ in DFA1 (blue) with the parameters defined in equation (4.11) and DFA3 (orange) with the parameters defined in equation (4.14).

of all three terms Λ_1, Ξ_1 and $\varkappa_1$ yield

$$\Lambda_1(s) = 60[s^2(g^2 - g)^2 - 3s(g^3 - g) + 2(g^4 + g^3 + g^2)],$$

$$\begin{aligned} \Xi_1(s) = \ & s^5(g - 1)^5(g + 1) + 15s^4 g(g - 1)^4 - 5s^3(g - 1)^3(1 - 7g - 7g^2 + g^3) \\ & -15s^2 g(g - 1)^2(1 - 10g + g^2) + 2s(g - 1)^3(2 - 17g - 17g^2 + 2g^3) \\ & -120g^2(1 + g + g^2), \end{aligned} \tag{4.11}$$

$$\varkappa_1(s) = 15(g - 1)^6(s^2 - s^4).$$

All three are polynomials in s. $\Lambda_1(s)$ is of order 2, $\Xi_1(s)$ of order 5 and for $\varkappa_1(s)$ it is 4.

In DFA3 the result was first presented in [D] (calculated by M. Höll). The components are $\Lambda_3(s)$ as a polynomial in s of order 6, $\Xi_3(s)$ as a polynomial of order 9 and $\varkappa_3(s)$ of order 8. The coefficients of fluctuation function derived from equation (2.59) are given

by

$$\begin{aligned}
\Lambda_3(s) =& 504s^6(g-1)^6g^2 - 7560s^5(-1+g)^5g^2(1+g) \\
&+ 6552s^4(g-1)^4g^2(7g^2+16g+7) - 7560s^3(g-1)^3g^2(19g^3+73g^2+73g+19) \\
&+ 1008s^2(g-1)^2g^2(242g^4+1357g^3+2382g^2+1357g+242) \\
&- 30240sg^2(-7-45g-99g^2+99g^4+45g^5+7g^6) \\
&+ 72576g^2(1+9g+45g^2+65g^3+45g^4+9g^5+g^6),
\end{aligned}$$
(4.12)

$$\begin{aligned}
\Xi_3(s) =& 2s^9(g-1)^9(g+1) + 63s^8(g-1)^8g - 12s^7(-1+g)^7(5-26g-26g^2+5g^3) \\
&- 126s^6(-1+g)^6g(7-30g+7g^2) + 42s^5(-1+g)^7(13-97g-97g^2+13g^3) \\
&+ 63s^4(-1+g)^4g(49-644g+710g^2-644g^3+49g^4) \\
&- 4s^3(-1+g)^7(410-3497g-3497g^2+410g^3) \\
&- 252s^2(-1+g)^2g(9-446g+143g^2-2292g^3+143g^4-446g^5+9g^6) \\
&+ 144s(-1+g)^7(8-71g-71g^2+8g^3) \\
&- 72576g^2(1+9g+45g^2+65g^3+45g^4+9g^5+g^6),
\end{aligned}$$
(4.13)

$$\varkappa_3(s) = 63(-1+g)^{11}(1+g)(s^8-14s^6+49s^4-36s^2).$$
(4.14)

The fluctuation functions of the AR(1) process for $g = 0.5$ in DFA1 and DFA3 are shown in figure 4.5. The scaling of both functions asymptotically goes to $1/2$. Both functions have a crossover and a steeper slope for small s. However, the asymptotic behavior is already visible for smaller s in DFA1 compared to DFA3. In this sense, the crossover appears later for higher orders of DFA.

4.1.4 Calculation rules

Real processes, in many cases, cannot be described by one simple model process. However, sometimes it is possible to approximate them as a combination of such models. Therefore, it is useful to know how transformations and superpositions change the fluctuation function of DFA.

Assuming the time series X is a superposition of two independently fluctuating processes Y and Z

$$X(t) = Y(t) + Z(t),$$
(4.15)

then the covariance of X is the superposition of the covariances of Y and Z, since

$$\begin{aligned}
\langle X(\tau+t)X(\tau)\rangle =& \big\langle \big(Y(\tau+t)+Z(\tau+t)\big)\big(Y(\tau)+Z(\tau)\big)\big\rangle \\
=& \big\langle Y(\tau+t)Y(\tau)\big\rangle + \big\langle Z(\tau+t)Z(\tau)\big\rangle \\
&+ \big(\big\langle Y(\tau+t)Z(\tau)\big\rangle + \big\langle Z(\tau+t)Y(\tau)\big\rangle\big).
\end{aligned}$$
(4.16)

The last two terms (in the brackets) are zero if the processes are independent. Therefore the correlation follows a superposition principle

$$\langle X(\tau+t)X(\tau)\rangle = \langle Y(\tau+t)Y(\tau)\rangle + \langle Z(\tau+t)Z(\tau)\rangle.$$
(4.17)
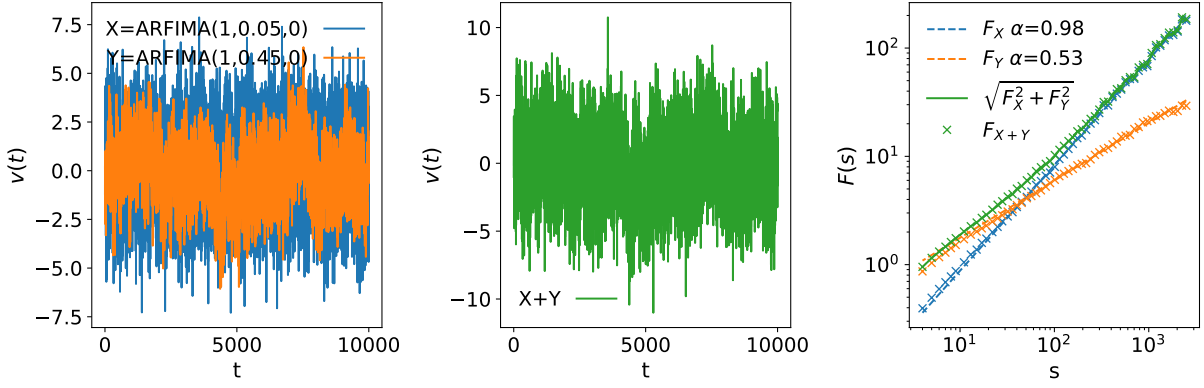
Figure 4.6

LEFT: Two time series: one generated by an ARFIMA(0,0.05,0) model and one by an ARFIMA(0,0.45,0) process. CENTER: Superposition of both time series. RIGHT: DFA for the two ARFIMA processes (orange and blue). The green dots are the DFA fluctuation function of the superimposed signal. The green line is the superposition of the two fluctuation functions according to equation (4.18). Green line and dots coincide.

Now this equation can be plugged into equation (2.59) with $C_{XX}(t) = \langle X(\tau+t)X(\tau)\rangle / \sigma_X^2$

$$
\begin{aligned}
F_X^2(s) &= \sigma_X^2 \sum_{t=1}^{s-1} C_{XX}(t) L_q(t,s) \\
&= \sum_{t=1}^{s-1} \langle X(\tau+t)X(\tau)\rangle L_q(t,s) \\
&= \sum_{t=1}^{s-1} \Big(\langle Y(\tau+t)Y(\tau)\rangle + \langle Z(\tau+t)Z(\tau)\rangle \Big) L_q(t,s) \\
&= F_Y^2(s) + F_Z^2(s)
\end{aligned}
\tag{4.18}
$$

So DFA inherits a superposition principle from the correlation function, which was already observed empirically [58]. Here it is tested numerically for the case of two ARFIMA(0, d, 0) processes in figure 4.6. The summation of independent processes looks relatively easy in DFA. This can be used for decomposing contributions to the fluctuation function, which will be discussed in chapter 5.

At this point, one more simple class of processes shall be presented that can be displayed as a combination of simpler models. The question is, what happens if the noise X, that is regarded as the input to a process Y, is already correlated. In general, the iteration relation in this case can be written as

$$
Y(t) = D\Big(X(t), Y(t-1), Y(t-2), ...\Big),
\tag{4.19}
$$
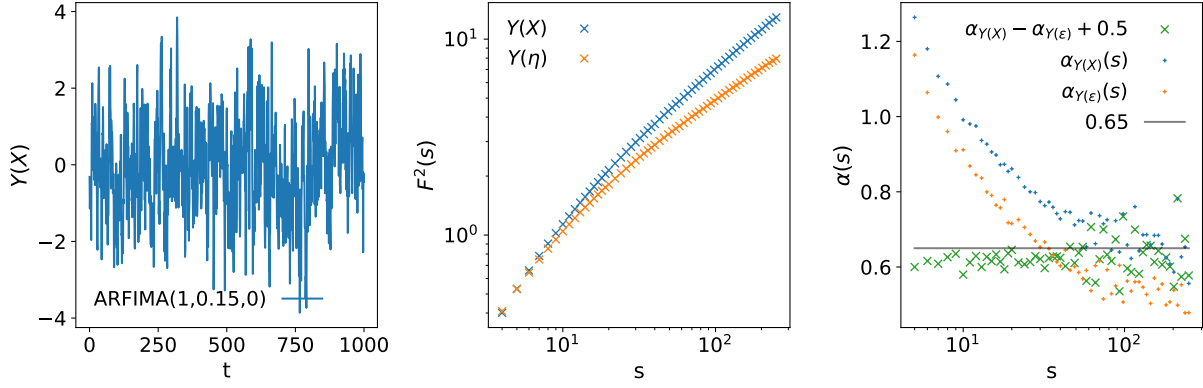
Figure 4.7

LEFT: Time series generated by ARFIMA(1,0.15,0) process with autoregressive parameter 0.5. CENTER: DFA fluctuation functions of the ARFIMA model $Y(X)$ with colored noise input and an AR(1) process $Y(\eta)$ with white noise input and identical AR-parameter. Both are averages of 1000 realizations. RIGHT: The incremental slopes of both fluctuation functions are indicated by the small dots. The blue crosses mark the difference between the slopes of both fluctuation functions plus $1/2$ according to equation (4.23). They are approximately equal to the expected value of $d + 1/2 = 0.65$.

where X is a correlated time series and D is some arbitrary function that determines the dynamics.

The problem can be exemplified by a system, where $X(t)$ and $Y(t)$ are both of the type of ARFIMA(0, d, 0), introduced in section 2.2.4. So X and Y are defined by the iteration relations

$$X(t, \epsilon(t)) = (1 - B)^{-d_X} \epsilon(t), \tag{4.20}$$

$$Y(t, X(t)) = (1 - B)^{-d_Y} X(t), \tag{4.21}$$

where ϵ is white noise and B is the backshift operator. The exponents d_X and d_Y are related to the DFA exponent α of the corresponding process via $d = \alpha - 1/2$. Now it is easy to see that the dynamics of $Y(t)$ could as well be directly defined as

$$Y(t, \epsilon(t)) = (1 - B)^{-d_Y}(1 - B)^{-d_X} \epsilon(t) = (1 - B)^{-(d_Y + d_X)} \epsilon(t). \tag{4.22}$$

So the combination of two ARFIMA(0, d, 0) processes yields again an ARFIMA(0, d, 0) process where d is the sum of the two parameters d of the two parameters d of the original processes.

The result for ARFIMA models has implications for other processes as well. As ARFIMA produces close to straight lines in the log-log plot of the fluctuation function, the fluctuation functions of other processes can be locally approximated by these straights. So in the environment of small changes Δs the implication is that local scaling

exponents add up if processes are combined in the sense of equation (4.19). Therefore, in general such processes yield exponents

$$\alpha_{Y(X)}(s) = \alpha_{X(\eta)}(s) + \alpha_{Y(\eta)}(s) - \frac{1}{2} \tag{4.23}$$

in DFA for a chosen value of s.

A local value $\alpha(s)$ can be calculated for a small segment Δs as

$$\alpha(s) \approx \frac{F(s + \Delta s) - F(s)}{\Delta s} \tag{4.24}$$

One example for a model that combines simple models is the previously mentioned ARFIMA(1, d, 0) process (see section 2.2.4). It can be described as an AR(1) process with ARFIMA(0, d, 0) as noise input instead of Gaussian white noise. Figure 4.7 is shows that equation (4.23) is valid for this model. The fluctuation function of ARFIMA(1, d, 0) (see time series in the left panel) is decomposed. In the central panel the fluctuation functions from the ARFIMA(1, d, 0) process and the fluctuation function of an AR(1) process with the same autoregressive parameter are shown. The values $\alpha(s)$ are calculated form both time series according to equation (4.24). In the right panel the scaling d is calculated from subtracting the values of $\alpha(s)$ for both fluctuation functions according to equation (4.23).

To summarize: while the additive composition of two signals leads to equation (4.18), inserting one process instead of white noise into another one leads to equation (4.23). As shown in figure 4.6 superposition can lead to convex crossovers of the fluctuation function in contrast to short range correlations, which lead to concave crossovers as the example of AR(1) shows.

4.2 Energy balance models for global mean temperatures

A typical application for DFA has always been atmospheric data, especially temperature time series [39, 47, 82, 126, 84]. Researchers have found the Joseph effect in local temperature data [39] as well as in average temperatures of large areas [126]. In this section, temperature is discussed in the most macroscopic from, namely the time series of global mean surface temperatures (GMST)s, i.e. averages of many measurements around the Earth. The question, whether or not temperature time series exhibit long range correlations, is important since long memory influences the error bars one has to take into account when modeling the data [84]. Therefore, it changes the significance of trends due to climate change, i.e. the significance of climate change itself. It also influences predictions of how much warming is expected in the future [47, 106].

It is known that a DFA exponent of $\alpha > 1/2$ for a short measurement time is no proof of long memory as an asymptotic property [82, 81]. The question which is addressed in this section is whether or not it is plausible that the obtained exponent is not the asymptotic

one. Therefore, the theoretical expression of the fluctuation function is employed, as well as a zero-dimensional climate model introduced by Hu and Fedorov [59], which is able to produce output very close to observed GMST data. The analysis was published in [D].

In this section real data from measurements around the world is used. Atmospheric data is made available by various sources. So it should be specified which dataset was used. Following [59], the atmospheric CO_2 concentration is given by Mauna Loa in situ measurements (`www.esrl.noaa.gov/gmd/ccgg/trends/`), which start in 1959 are combined with ice core reconstructions from Law Dome DE08 and DE08-2 (`cdiac.ornl.gov/ftp/trends/co2/lawdome.smoothed.yr20`) before 1959. The volcanic forcing is proportional to the stratospheric aerosol optical depth (SAOD) data set from NASA GISS [110]. Unlike Hu and Fedorov here temperature measurements from the Hadley Center Met Office web site (`crudata.uea.ac.uk/cru/data/temperature/`) are used, because the record is longer compared to the dataset from the Goddard Institute of Space Studies (GISS). The length of the dataset is a crucial criterion since long range correlations are an asymptotic property.

4.2.1 A zero dimensional energy balance model

The most simple model for global mean temperatures is a zero-dimensional energy balance model [51, 108]. Zero dimensional here means that no spacial dimensions are considered and the earth is considered as a point. It is written as a differential equation for GMST, which includes deterministic and/or stochastic forcings.

Such a model can be motivated physically. The following derivation is well known and can be found e.g. in the appendix of [105]. The energy balance of the earth is determined by the incoming radiation from the sun I and the outgoing radiation, which is given by the Stefan-Boltzmann law. It affects the ocean surface temperature T_s depending on the heat capacity C_S

$$C_S \frac{\mathrm{d}T_S}{\mathrm{d}t} = -\sigma_B T_A^4 + I. \tag{4.25}$$

Here the variables in the Stefan-Boltzmann law are the Temperature of the lowest atmospheric layer from which radiation can escape T_A and the Stefan-Boltzmann constant σ_B. Now assuming that the temperature T_A and T_S are linearly related to each other and to the global mean surface Temperature $\hat{T} = T_A/a = T_S/b$ and also rewriting the solar irradiance as a function of the flux density $J = I/c$ and the albedo constant c, one obtains

$$C_S \frac{\mathrm{d}(b\hat{T})}{\mathrm{d}t} = -\sigma_B a^4 \hat{T}^4 + cJ. \tag{4.26}$$

So a first order differential equation can be formulated for the temperature $\hat{T}$. When looking at the fluctuations the quantity of interest is not the absolute temperature $\hat{T}$ but the temperature change T with respect to some equilibrium level T_0, i.e. the pre-industrial temperature. Therefore one can use the ansatz $\hat{T} = T_0 + T$ in equation (4.26)

and linearize it using a Taylor expansion. The remaining quantities are expanded like

$$J = J_0 + \Delta J$$
$$a = a_0 + \frac{\partial a}{\partial T}T + \Delta a$$
$$b = b_0 + \frac{\partial b}{\partial T}T + \Delta b$$
$$c = c_0 + \frac{\partial c}{\partial T}T + \Delta c. \tag{4.27}$$

In addition, the constant C_S can be expressed as $C_S = C/b_0$. The terms in the first order depend on $T, J_0, a_0, \Delta a$ and Δc. They are substituted by the climate sensitivity parameter S_{eq}. The change in T can also be due to a forcing $K(t)$, that might be due to changes in albedo or greenhouse gas concentration

$$C\frac{dT(t)}{dt} = -\left(\frac{1}{4\sigma_B a_0^4 T_0^3}\left(1 + \frac{T_0}{a_0}\frac{\partial a}{\partial T}\right) - \frac{\partial c}{\partial T}J_0\right)T(t) + \left(\Delta c J_0 + c_0\Delta J - \frac{4\sigma_B c_0 J_0 \Delta a}{a_0}\right)$$
$$= -\frac{1}{S_{eq}}T(t) + K(t). \tag{4.28}$$

Equation (4.28) is a linear differential equation for T. Spatial differences in T are not resolved in this zero-dimensional model. The climate sensitivity is discussed, for example, in [108, 113]. C can be approximated by (τ/S_{eq}) with the relaxation time τ. T is called the global mean surface temperature anomaly.

The forcing K drives T away from its equilibrium value $T = 0$. It might include more terms than the solar irradiance and the greenhouse gas concentration, which are present in the above derivation. Important drivers of climate are also human-made aerosols and volcanic activity [51].

As mentioned above, there where several investigations on the autocorrelation properties for long times in temperature data. Long range correlations, i.e. the Joseph effect, were found in many time series by several researchers using several methods [39, 107, 91]. The ordinary differential equation (4.28) does not seem to be suitable to reproduce these findings. Therefore, Rypdal and Rypdal [108] propose to model GMST with a fractional differential equation instead, which they also motivate from the energy balance. Just as for ARFIMA (see section 2.2.4) fractional differentiation in an energy balance model is able to generate long range correlations.

The question that should be answered here is about whether or not the explicit modeling of long range correlations in the spirit of Rypdal and Rypdal is really necessary in order to reproduce the temperature variability. The question pops up, since many models, including very recent ones, use traditional first order differential equations [52, 59], and at first sight seem to reproduce the fluctuations pretty accurately.

A specific example of such a model was introduced by Hu and Fedorov [59]. They only take yearly averaged data into account. The model is not stochastic, but instead includes an index of the El Niño Southern Oscillation (ENSO) as additional forcing.

Apparently, it well approximates the global fluctuations on the timescale of a few years. The energy balance equation reads

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -\frac{T}{\tau} + a \log\left(\frac{\mathrm{CO}_2}{\mathrm{CO}_{2,r}}\right) + bT_{Nino} + c\mathrm{SAOD} + d. \qquad (4.29)$$

Here the amplitudes a, b, c and d are free parameters to be fitted from data. The model does not include solar irradiance since its effect is small. It further assumes the effects of changes in aerosol concentration, which lead to cooling, and changes of greenhouse gasses other then CO_2, which lead to warming, are approximately equal and therefore cancel each other. The CO_2 concentration is divided by the reference concentration $\mathrm{CO}_{2,r}$, which was chosen at 320 ppm. This represents a pre-industrial level. The ENSO forcing T_{Nino} is calculated from the sea surface temperature (SST) anomalies HADSST3 [69]. It is not one of the standard indices, but averages over a larger region of the southern pacific. This measure for ENSO was proposed in [124]. T_{Nino} is calculated by first averaging the sea surface temperature between $160°\mathrm{E} - 90°\mathrm{W}$ and $5°\mathrm{S} - 5°\mathrm{N}$ in the tropical pacific. Then the global mean SST time series is subtracted from each value. Due to climate change, this data still exhibits a small trend, which is removed by subtracting a linear fit. The yearly averaging of all measurements was done for July to June in order to not cut the ENSO signal at the peak [59], as El Niňo typically starts around Christmas.

So equation (4.29) is a simple model describing GMST. The temperature depends on several forcings, which are measured time series, themselves. It should now in principle be possible to fit the weights of the various forcing components, i.e. the parameters a, b, c, d, and, in addition, the relaxation time τ. In reality, fitting is very difficult, since the model has five free parameters and the datasets consist only of around 150 common years of recordings. Therefore, multiple combinations of parameters might lead to a similarly small error.

At this point, one might think of solutions to increase the quality of the fit, however, for the purpose of this chapter, namely to reproduce long range correlations in a simple model, it is irrelevant how exactly the model looks like. The focus is not on finding the correct weights of the forcing terms accurately, but it is sufficient to provide one possible model which reproduces the temperature variability.

Since there is no better information available the relaxation time τ is chosen as the same value that was suggested in [59], so $\tau = 2$ years. This is a short time compared to other estimates that were done e.g. in [113, 108].

The remaining parameters are fitted in such a way that the discrepancy of the model output and the observed GMST signal at every time step (year) is minimal, but the parameters at the same time fulfill the constraint that those, which are most relevant to this work, namely τ and b, are close to the values obtained in [59]. The result is shown in figure 4.8. Newer data points contribute to the fit with a higher weight than older values. The HADCRUT4 GMST data [92] is in general fitted reasonably well by the model data. In the early years, in the nineteenth century, the deviations are much larger than in recent years. This is plausible since over time the measurement accuracy
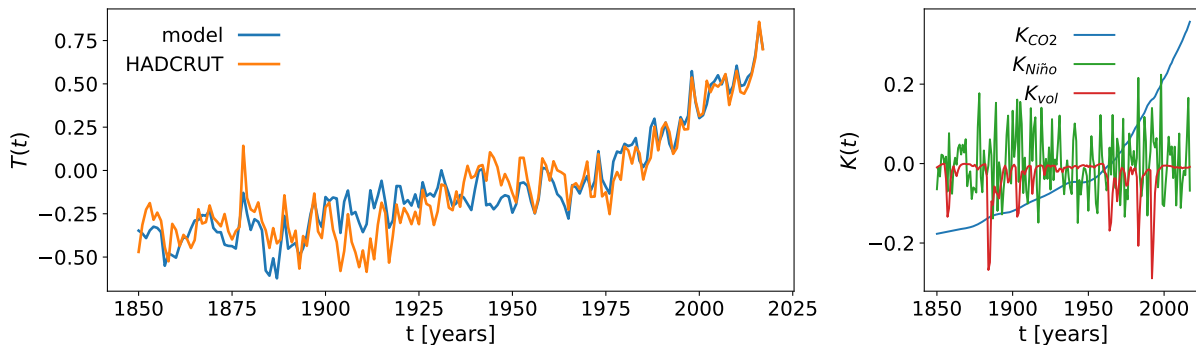
Figure 4.8

LEFT: Model output of equation (4.29) and HADCRUT4 data. RIGHT: Forcings for the model, according to the equations (4.29), and (4.30).

has strongly increased due to the installation of many additional measurement stations. The chosen parameters are

$$
\begin{array}{lll}
a = 1.51 & c = -2.40 & \tau = 2.00 \\
b = 0.122 & d = -0.0035 &
\end{array}
. \tag{4.30}
$$

The right panel of figure 4.8 provides a plot of the forcings K_{CO2}, K_{Nino} and K_{vol}. They are given by $\log(CO_2/CO_{2,r})$, T_{Nino}, and SAOD, multiplied by the corresponding parameters given in equation (4.30). As mentioned before, the fitted parameters are not reliable, because there is not enough data available for a perfect model fit. However, this is not a big concern, since it is sufficient to conclude that equation (4.29) with the given parameters is obviously one out of several possible models, that is able to approximate the available GMST data. The plot of both time series is shown in figure 4.8 in the left panel. It raises confidence that this is the case. In the next section, it shall be analyzed whether or not the model is able to reproduce and explain the measured long range correlations in the data.

4.2.2 DFA for data and the model

The output of the energy balance model, the GMST observations, and various components of the forcing in this model are analyzed by DFA. It is shown that on time windows of typical observations, DFA shows the same signatures for observed data and model output, even though the model can not generate long range correlations in the true sense.

Past studies have shown that DFA3 is sufficient for detrending the temperature increase due to climate change and DFA4 or DFA5 are no further improvements [39]. This can easily be reproduced. Therefore in this section DFA3 is employed for GMST data and stationary forcing data, i.e. ENSO and SAOD, DFA1 is used because it is less data consuming.
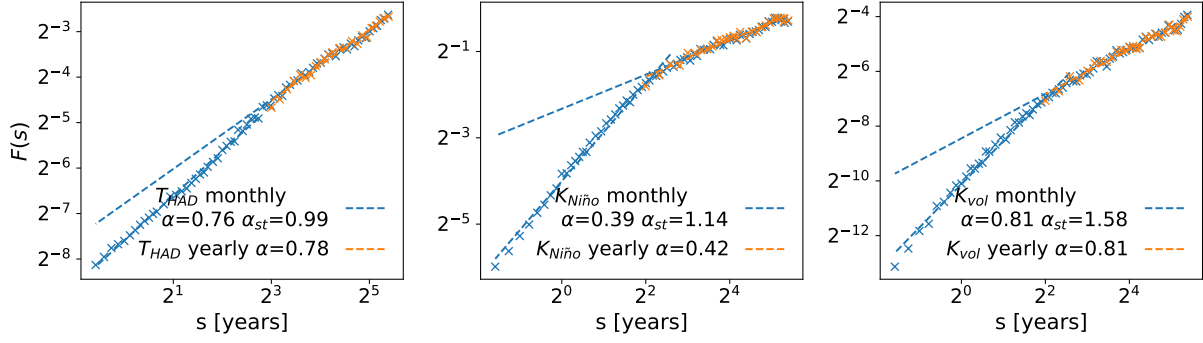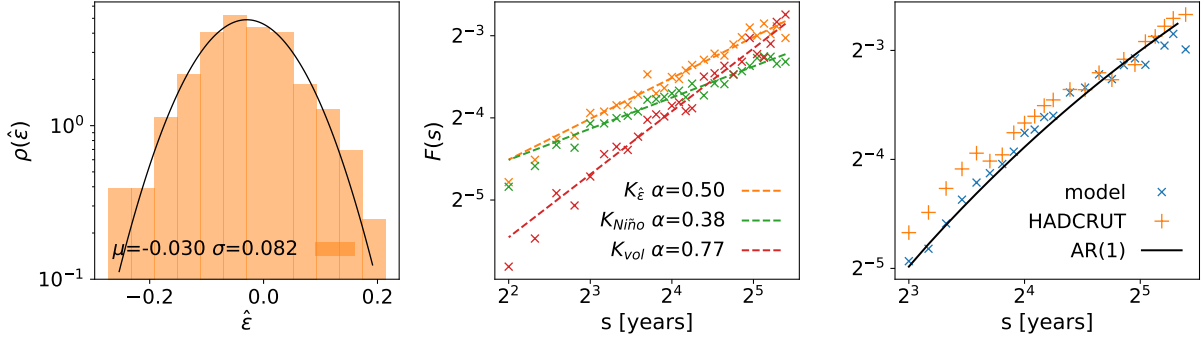
Figure 4.9

LEFT: DFA3 for monthly and yearly GMST data; CENTER: DFA1 for monthly and yearly K_{Nino} and (RIGHT:) F_{vol}. The fit yields $\alpha = 0.76$ to 0.78 for GMST, $\alpha = 0.40$ to 0.41 for K_{Nino} and $\alpha = 0.76$ to 0.78 for K_{vul} on the timescale of years. It is significantly higher for shorter times in the monthly recorded datasets with yearly cycle removed.

The GMST record is known to show long range correlations [107]. Its values are averages from measurements around the globe, which do not all scale in the same way. While for individual stations on land the measured exponents α are mostly smaller or equal to 0.65, on the oceans they are larger up to almost $\alpha = 1$ [39]. However, it is not clear if all of these measured exponents represent the asymptotic behavior. If they did, the asymptotic exponent of the averages GMST would be equal to the largest local exponent due to the superposition principle of DFA (4.18)

$$F^2_{\Sigma_i x_i}(s) = \sum_i F^2_{x_i}(s) > \max_i (F^2_{x_i}(s)) \quad \forall s. \tag{4.31}$$

The fluctuation function of the temperature time series $T(t)$ is shown in figure 4.9 (left panel). The exponent obtained from HADCRUT4 data is 0.80 on timescales of several years. The analysis is shown for both yearly and monthly data. This helps to get an idea about uncertainties, as the fits differ from each other a little bit. If s for monthly data is divided by 12, there is no difference between the monthly recorded data $T(t_m)$ and yearly averaged data $T(t) = \lfloor T(t_m) \rfloor_{12}$ for identical time ranges of s. This means results are consistent with each other and the asymptotic scaling is not changed by averaging. The short timescales, visible in the analysis of the monthly data, are not of interest here since they are not covered by model (4.29). Nevertheless, one might remark that the exponent is significantly higher ($\alpha_{st} \approx 1$) for small s. This strong short range correlation is also apparent in the results of [126].

In figure 4.9 in the central and right panel, fluctuation functions for SAOD and T_{Nino} data are shown. For these datasets, the behavior on the timescale of years differs even more from the small s behavior. For the ENSO signal, the apparent asymptotic exponent is $\alpha = 0.42$, indicating anti-persistence, while for short times the correlations are clearly positive. SAOD exhibits long range correlations with $\alpha = 0.81$ for large s.

Figure 4.10

LEFT: Probability distribution of the fluctuating forcing term $\hat{\xi}(t)$ in a logarithmic scale. A Gaussian distribution drawn as black line is fitted. Its mean value is -0.030 and the standard deviation is 0.082. CENTER: DFA for the both constituents as well as the full fluctuating forcing $\hat{\xi}$. While K_{Nino} appears anti-persistent and K_{vol} long range correlated in the considered time range, the full signal appears uncorrelated. RIGHT: DFA fluctuation function for the yearly GMST data and the model (4.29) compared to the theoretical fluctuation function of AR(1).

It is known since the observations of Maraun et al. [82] that fluctuations in yearly GMST data match quite well with an AR(1) model. By combining theoretical knowledge about DFA with the Hu & Fedorov model from equation (4.29), one can physically motivate the interpretation of the DFA result as short range correlations.

In a first step, the model can be shown to be by a kind of autoregressive process by separating fluctuating and non-fluctuating components of the forcing. It can be written as

$$T(t) = (1 - \tau^{-1})T(t-1) + \hat{\xi}(t) + a \log\left(\frac{\mathrm{CO}_2(t)}{\mathrm{CO}_{2,r}}\right) + d, \qquad (4.32)$$

with a total (deterministic) fluctuating component

$$\hat{\xi}(t) = K_{Nino}(t) + K_{vol}(t). \qquad (4.33)$$

The CO_2 concentration adds a trend to the signal that should not contribute to the result in DFA3.

In figure 4.10 in the left and central panel, the fluctuating component $\hat{\xi}$ is analyzed. In the left panel, one can see that the distribution is Gaussian in a good approximation. This seems to be mainly due to the distribution of the ENSO forcing, which is close to Gaussian itself. The distribution of K_{vol} is asymmetric.

In the central panel, the fluctuation functions of both components and the sum are presented. For short times, up to ~ 4 years, the exponent α of the fluctuation functions $F_{K_{Nino}}$ and $F_{K_{vol}}$ are very high. However, white noise is a good approximation for the combined signal $F_{\hat{\xi}}$ on longer timescales of years. Here DFA1 yields no correlations. So going back to equation (4.29), with $\hat{\xi}$ replaced by white noise, the model looks like an

83

autoregressive process with an additional trend due to the CO_2 forcing. In conclusion, the fluctuations in the model are statistically similar to an AR(1) process.

Now what does this mean for the result in figure 4.9, which shows the DFA of the GMST record. The fluctuation function of the model was obtained numerically on the model output using the observed driving forces. The theory curve for the corresponding AR(1) process is given in equation (4.10). The measured fluctuation function of GMST can be compared to both. The results are shown in figure 4.10 in the right panel.

Indeed the DFA result of the model excellently matches the theory for AR(1). The measured data has a less steep slope, i.e. a bit stronger fluctuations for short times. So the model actually over-represents the observed effect of long range correlations. For the longest times available all three fluctuation functions coincide. It can be concluded that equation (4.29) is able to reproduce the DFA result even though it does not contain dynamical long range correlations.

### 4.2.3 Implications for longer time series

The scaling of the fluctuation function for GMST for timescales around a decade can be well explained by a model with short range correlations. Such a model can be physically motivated by an energy balance equation. The fluctuating forcing terms in this model are the El Niño Southern Oscillation and the stratospheric aerosol optical depth, which is an indicator of volcanic activity. Both constituents can clearly not be described as uncorrelated white noise. However, depending on the choice of parameters, the superposition might look uncorrelated for the timescale of years. The climate response was estimated with an AR-parameter of 0.5.

For the accessible range of s in the fluctuation function, the short range correlated model can not be distinguished from a long range correlated model. The question that will be addressed in this section reads: How long does a time series of the proposed model need to be in order be clearly identifiable as short range correlated?

It is known that the AR(1) process being short range correlated scaled like $s^{0.5}$ for $s \to \infty$ in DFA. In figure 4.11 the full fluctuation function of the obtained simplified model (4.29) for s up to 1000 years is shown. Its scaling is fitted for different ranges of s. Only for $s \approx 1000$ the fitted exponent is close to 0.5 for the AR parameter $g = 0.5$ corresponding to equation(4.29).

So despite the correlation time of around two years is moderate, the asymptotic behavior is only accessible with DFA3 after very long observation times. This makes it very difficult to distinguish short range correlations from long range correlations. Presumably, the underlying problem will be present in all measures for long range correlations one might think of. Locally a power law curve can always be approximated by an exponential and vice versa. However, in DFA, a systematic effect adds to this, which shifts crossovers to longer times [65, 55]. The effect gets stronger the higher the order q of DFA is. So it is the price one has to pay for the detrending of the time series.

For the discrete wavelet transform the onset of the asymptotic behavior of this method for the AR(1) model seems to appear earlier at scales $s \approx 200y$, but at the expense of much stronger statistical fluctuations, i.e. longer time series of length t are needed in
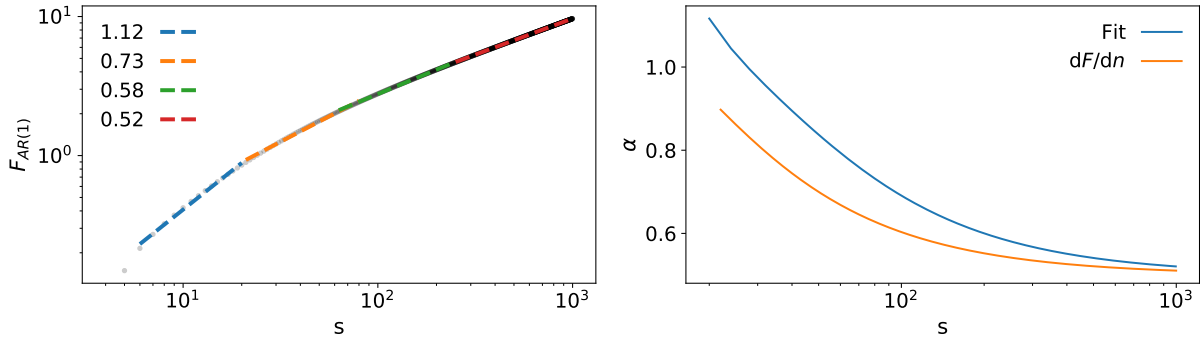
Figure 4.11

Illustration of DFA applied to AR(1): a range of α-values can be obtained depending on the range s. LEFT: Theory curve (4.10) for the fluctuation function of the AR(1) process for DFA3. Several linear fits were performed for different ranges of s. α decreases as s increases. Its values for $s = 20, 80, 250, 1000$ are printed in the picture. RIGHT: results for fits at different s and the logarithmic derivative of the fluctuation function. Both converge to 0.5 for large s.

order to calculate values for the same s. It was not part of this investigation to compare both methods. So wavelet analysis may or may not perform better than DFA, however, the recorded time series for GMST are clearly too short for both DFA and Wavelet analysis to observe the crossover. The fluctuation functions of the proposed model, the HADCRUT data and the AR(1) process all yield long range correlations and are difficult to distinguish.

The investigation strengthens the point of Maraun et al. [82], who already compared temperature fluctuations to the AR(1) model, and also Mann [81] who only looks at stochastic components that are added to measured components of the forcing. However, the investigations do not disprove the presence of long range correlations. They might be present in climate response [108] or in the climate forcing. Scientists have studied the general circulation climate models for very long times, e.g. 1000 years of simulated local temperatures exhibit long range correlations [104]. The question of long range correlations in temperatures will pop up again in chapter 5 in a different setting, where oscillatory modes in the climate dynamics will be investigated.

4.3 Direct fit of linear response models

The previous sections summarized the properties of DFA and showed one application of how the theory helps to understand the results of simple models. However, the real analysis was performed without using DFA. The data was fitted directly with the energy balance model. While traditional usage of DFA means calculating the Joseph exponent, for short range correlated models like the one of Hu and Fedorov [59] this is not necessary.

In this section, DFA is finally used as an analysis tool that can be applied for inferring information and constructing data models in the absence of long range correlations.

Now for this purpose, the whole fluctuation function of a theoretical model is taken into account and fitted to the fluctuation function of the data. This idea was published in [F]. In order to keep things simple the theoretical model, for now, will always be AR(1), which describes linear response.

Applications to a toy model and to real data are presented. The data for global mean surface temperatures is again taken from the Hadley Center Met Office web site (`crudata.uea.ac.uk/cru/data/temperature/`) like in the previous section. In addition, applications to pressure time series are presented, which are provided by the ECA&D project [120], publicly available online at `http://www.ecad.eu`.

4.3.1 Fitting algorithm

The theoretical fluctuation function of AR(1) only has two free parameters, g and σ (see Eq 4.10). Therefore, fitting is usually possible even for short datasets without caring too much about the stability of the fitting routine. However, stability can be improved even more when both parameters are obtained separately. Since there will be more parameters to fit in the next chapter, a solid fitting procedure is crucial. In each case fitting is done by minimizing the variance

$$\mathrm{var}\left[\log\left(\frac{F_{data}(s)}{F_{model}(s)}\right)\right].\tag{4.34}$$

In this way g is obtained. The standard deviation σ is not fitted but rather calculated as the offset of F_{model} with respect to F_{data}

$$\sigma = \exp\left(\sum_{s\in S}\log\left(\frac{F_{data}(s)}{F_{model}(s)}\right)\right).\tag{4.35}$$

This algorithm proposed in [G] has turned out to be stable and give good results.

4.3.2 Relaxation times

In order to understand the meaning of AR(1) fits in DFA, it is useful to look at a modified version that illustrates the response of the system to perturbations more clearly. It was also presented in [F]. Instead of the AR(1) iteration equation (2.21) consider an idealized random process with spikes. The spike magnitude distribution $\rho(\xi)$ is an exponential distribution, so all moments exist. The waiting times $W(r)$ are exponentially distributed, too. In other words, the probability of a spike each time step is identical. The third property is the decay rate. It is chosen as $\gamma = 0.2$ for all spikes.

According to definition (2.21), this is an AR(1) process with Poissonian noise η which is not independent identically Gaussian distributed, yet uncorrelated. The AR parameter is $a = e^{-\gamma}$. In the left panel of figure 4.12, one realization of the process is shown. The right panel shows the fluctuation function F of a time series of length 3000. The theoretical AR(1) fluctuation function is fitted by the method presented in section 4.3.1 to the measured fluctuation function. All timescales are equally taken into account for
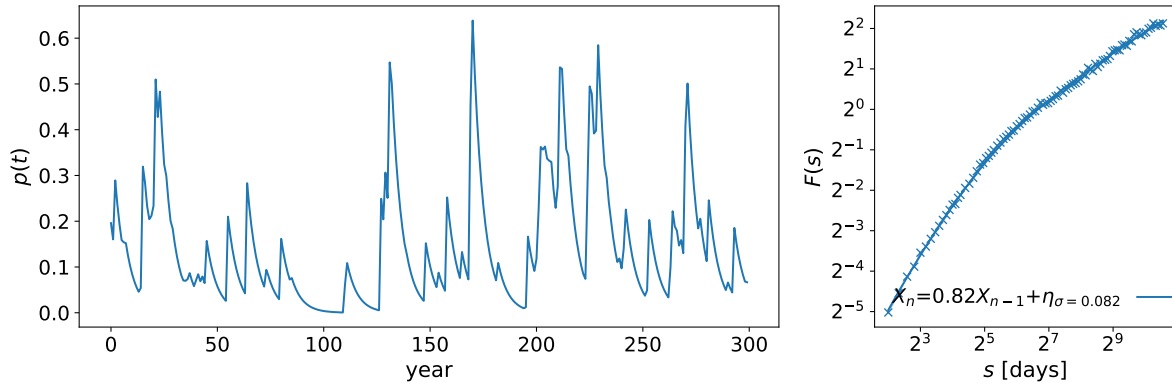
Figure 4.12

A spiking process with $\rho(\xi) = \exp(-10\xi)$ and $W(r) = \exp(-0.3r)$. Each spike decays exponentially with relaxation time 5. LEFT: time series of the process. RIGHT: DFA-1 fluctuation function of a time series (dots) with length 3000 and a theoretical AR(1) fluctuation function, fitted by the method of least squares. For 1000 trials (not in the picture) the obtained AR parameter is $a = 0.816 \pm 0.022$, which yields a relaxation time of $\tau = 4.92$. The real relaxation time $\tau = 5$ corresponds to $a = 0.819$. The result indicates a small bias and a larger variance.

the fit. The AR(1) coefficient obtained by the fit is, in a very good approximation, $a \approx e^{-\gamma}$. As expected, the spike magnitude and the spike frequency only affect the values of F by a global pre-factor.

Thus the method of fitting DFA fluctuation functions uncovers relaxation times in noisy time series. The method has advantages and disadvantages that originate from the properties of DFA. The fluctuation function is a smoothed measure that makes fitting easier and more robust compared to traditional measures like the correlation function or the power spectrum. This property, on the other hand, means that some features might not be visible anymore. For the generation of data model, this can even be an advantage if interpretations are formulated with care.

The biggest advantage compared to the correlation function is the accessibility of long timescales up to one-fourth of the measurement time. All timescales are equally taken into account due to the logarithmic scale. This leads to better models for intermediate and long times than a short-time-fit of the correlation function yields. The results can also immediately be validated visually and bad fits can be identified and discarded. This validation, as well as the method, is independent of the probability distribution and only takes the correlations into account.

4.3.3 Direct fit of climate response

The same method of fitting fluctuation functions can be applied to real data. In section 4.2, a simple model for global mean surface temperature was discussed. Yearly averaged GMST can be approximated by an AR(1) model. There are two main drivers of
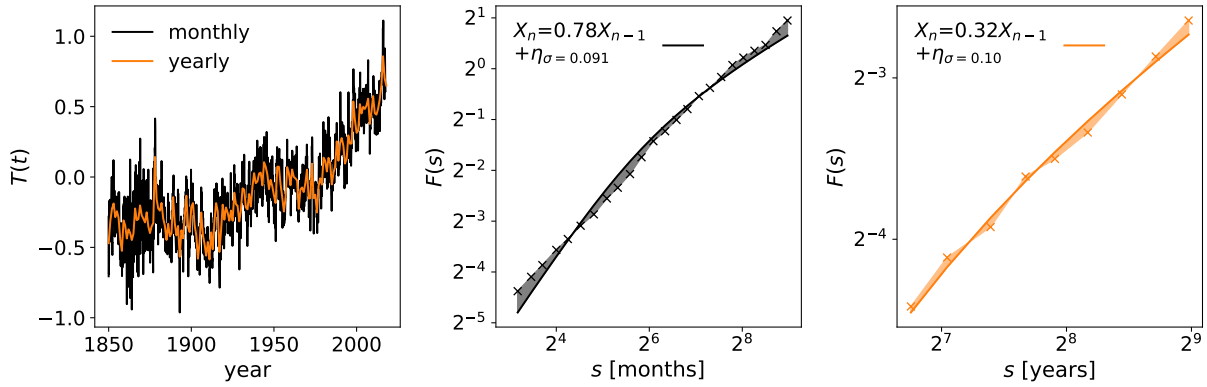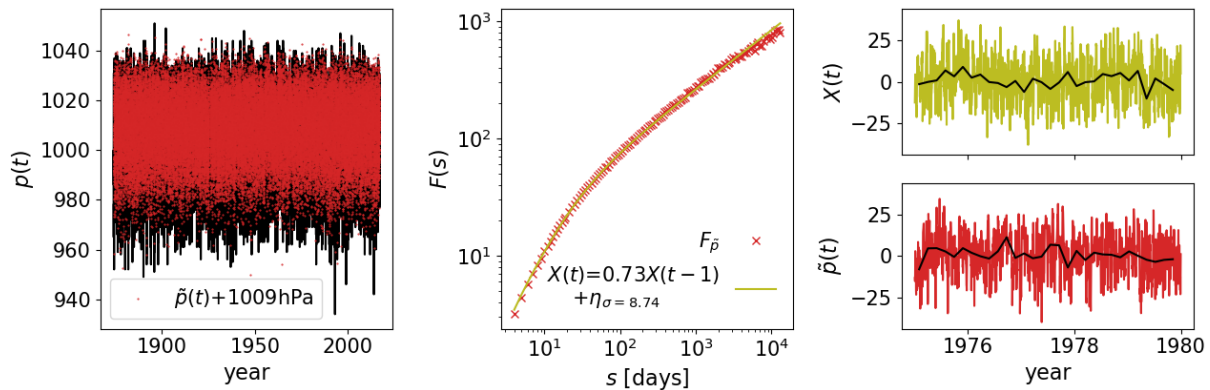
Figure 4.13

LEFT: GMST data since 1850 monthly averaged (black) and yearly averaged (orange). RIGHT: DFA-3 of both time series (fluctuation function of the yearly data has an offset in order to not lie on the monthly curve). Both are fitted by the AR(1) model. For yearly data, the parameter is $a = 0.32$, which yields a relaxation time of $\tau = 10.2m = 0.85y$. For monthly data, the parameter is $a = 0.78$, which yields a relaxation time of $\tau = 4.25m$. As explained in the text, we interpret this contradiction as a failure of the AR(1) model for monthly data, as visible by the systematic deviations of the fit from the empirical values (black line and black crosses).

global temperature fluctuations. These are volcanic eruptions and the El Niño Southern Oscillation. Yearly averaged ENSO is slightly anti-persistent for a few years, while the volcanic forcing is correlated with $\alpha > 0.5$ on this timescale. When both are super-imposed, the result in DFA looks like uncorrelated noise with $\alpha \approx 0.5$ for the relevant timescales. In section 4.2 these forcings were taken as inputs for a linear response AR(1) model where the AR-parameter g, and therefore also the relaxation time, were given by the result by Hu and Fedorov [59] to be $g = 0.5$. By fitting the fluctuation function a better result can be obtained here.

The obtained fluctuation function $F(s)$ of GMST is fitted with the theoretical fluctuation function of AR(1) in DFA3. The fit is shown in figure 4.13. It yields an AR-parameter of $g = 0.32$, which is significantly lower than the one used in [59] ($g = 0.5$). However, the goodness of the fit implies, that such a linear response model with relaxation time $\tau = 10.2m$ is a better model for climate variability of the timescale of years compared to the one shown in figure 4.10. When performing DFA on monthly data and fitting again the fluctuation function of AR(1), the obtained result is $g = 0.78$ in monthly time units. This yields a correlation time, which is even shorter than half a year and much shorter than the one obtained for yearly data. Where does this apparent contradiction originate? A closer look at figure 4.13 reveals that the fit for monthly data is bad, i.e. there are significant and systematic deviations between the two curves. The data is not well described by AR(1). This is because of intraseasonal climate variability, which can not be described by white noise and thus affects the correlations of the linear response model. AR(1) models should not be fitted to data blindly, but the plot

Figure 4.14

LEFT: Sea level pressure from Torshavn. The black lines mark the recordings while the anomalies are plotted as red dots. CENTER: DFA of sea level pressure anomalies (red) fitted by theoretical fluctuation functions of AR(1) (yellow). RIGHT: 5 years of the Torshavn anomaly time series (red) compared to an equally long time series of a realization of the fitted AR(1) model (yellow). The black lines indicate 60-days means $\lfloor \tilde{p}(t) \rfloor_{60}$.

of the fluctuation functions should be used for validation. Systematic deviations are warnings. The issues are similar to the problems of interpreting traditional DFA with fitting straight lines to fluctuation functions.

4.3.4 Sea level pressure

Fitting the GMST fluctuation function with AR(1) can be a useful tool for generating a simple data model for the fluctuations. As the failed attempt for monthly measurement already shows, this model is usually a huge simplification and very far from reality, at least for short times. This is because in reality several processes and atmospheric cycles are visible in temperature data as discussed in chapter 1. However, it turns out that for sea level pressure these patterns are usually less visible and the data is close to plain chaotic or AR(1) data. In this section, this observation will be exemplified with two time series of European sea level pressure measurements. The two stations are located at Torshavn, Faroe Islands and Zagreb-Gric, Croatia. These are two of the longest records of local sea level pressure time series available. The analysis was previously presented in [H].

The two datasets under investigation are plotted in the left panels of figure 4.14 and 4.15. Sea level pressure in Europe exhibits seasonality that, in contrast to temperature seasonality, is not mainly described by periodic oscillations of the first moment. In fact, the strongest periodic trend is present in the second moment. In order to remove this non-stationarity, one can use the algorithm presented in section 2.2.2. This is done here. The resulting anomalies $\tilde{p}(t)$ are shown in the figures.

In the central panels of figure 4.14 and 4.15, the DFA fits of the AR(1) model to the
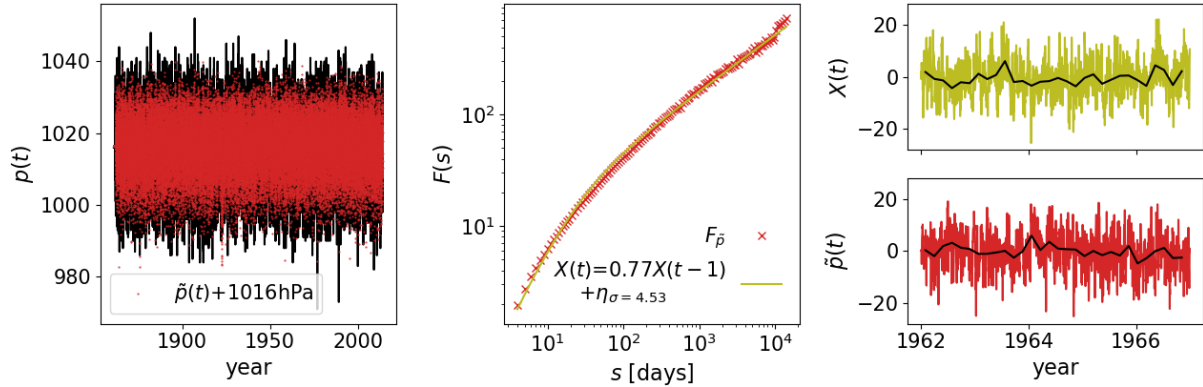
Figure 4.15
LEFT: Sea level pressure from Zagreb-Gric. The black lines mark the recordings while the anomalies are plotted as red dots. CENTER: DFA of sea level pressure anomalies (red) fitted by theoretical fluctuation functions of AR(1) (yellow). RIGHT: 5 years of the Zagreb anomaly time series (red) compared to an equally long time series of a realization of the fitted AR(1) model (yellow). The black lines indicate 60-days means $\lfloor \tilde{p}(t) \rfloor_{60}$.

two fluctuation functions are shown. A closer look at the figures shows that the fit for Torshavn is very accurate and the crossover is described well. For Zagreb, however, there are small systematic deviations, especially around the crossover. So for Torshavn the fitted relaxation time of 3.21 days seams to reflect the dynamics of the physical system, while for Zagreb the dynamics is more complicated. The calculated correlation time of 3.74 days can only be regarded as an approximate quantity.

Despite some small but systematic deviations in the fit for Zagreb, the obtained AR(1) process still is a nice and simple data model that describes the fluctuations on all considered timescales approximately. The meaning of the fitted results as data models are clarified by the plots in the right panels of figure 4.14 and 4.15. Here the measured time series is compared to a realization of the obtained models. The fluctuations of the time series $\tilde{p}$ and, at the same time, for the 60 days mean $\lfloor \tilde{p}(t) \rfloor_{60}$, are well captured by the AR(1) model for both datasets.

In this chapter, it was shown that actually proving long range correlations is an impossible task and short range correlations can be misinterpreted for long range correlations, even if measurements were done for a long time. In the final sections, however, it was also shown, that DFA is not useless despite its limitations. If one does not see the effects of short range correlations in DFA as undesired artifacts but instead concentrates the investigations on these properties, one can infer characteristic timescales and obtain data models that correctly describe the fluctuations on a long time range. In the next chapter, some additional features of the method of DFA fitting are introduced and more applications are presented.

5 Fluctuations on multiple timescales

A linear response model is not satisfying for most measured time series. Therefore, some extensions to the previously introduced method are necessary. At least, the method should be able to deal with oscillations and fluctuations on more than one characteristic timescale.

The effect of oscillatory dynamics is discussed in section 5.1.1. How to handle fluctuations on multiple timescales is explained in section 5.1.3. In the remaining sections, real-world examples are discussed. The topics are climate indices, measured temperature time series and power grid frequency fluctuations.

Since the superposition principle of the squared DFA fluctuation functions (4.18) plays a major role in this chapter, the squared fluctuation function $F^2(s)$ is always considered instead of $F(s)$.

5.1 Identifying timescales

In order to be useful for generating data models, the method of fitting fluctuation functions has to be improved. Firstly, it should be able to deal with oscillating dynamics or other more complicated response patterns. This can be covered with AR(2) models. They are discussed in section 5.1.1. Secondly, complex systems like the atmosphere fluctuate on multiple timescales. For example, the seasonal cycle and the El Niño Southern oscillation are superimposed and, in fact, also influence each other [118]. It turns out that such fluctuations on multiple timescales become rather simple in DFA due to the superposition principle, introduced in section 4.1.4.

For the analysis in section 5.1.2, data for global mean surface temperatures (HAD-CRUT4) and one ENSO measure, based on HADSST3, is taken from the Hadley Center Met Office web site (`crudata.uea.ac.uk/cru/data/temperature/`) just like in the previous chapter. The analysis was published in [F]. Here one additional ENSO measure is considered. It is the southern oscillation index, which is also made available by the climate research unit and can be downloaded at `https://crudata.uea.ac.uk/cru/data/soi/`. In section 5.1.4, Indonesian sea level pressure data is analyzed following [H], it is available at `http://sacad.database.bmkg.go.id`, as part of the ICA&D project [125].

5.1.1 The autoregressive model of order two

In the previous chapter, short range correlated processes were modeled by Markovian linear response. Due to its simplicity, the applications of the model are limited. Therefore, in addition to the autoregressive model of order one, it is useful to also consider
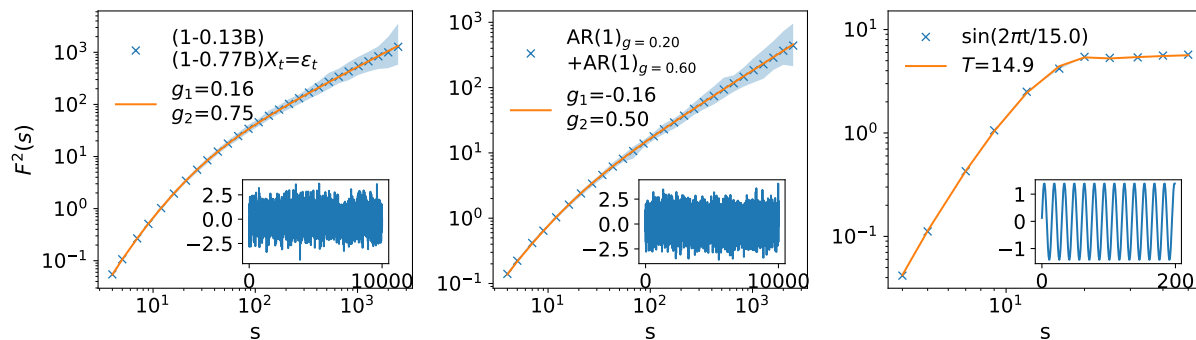
Figure 5.1

Fitting fluctuations function with AR(2). Trajectories are shown in the small panels. The fluctuation functions are averaged over 300 realizations. The blue stripes around $F^2(s)$ represent the error bars. LEFT: A process is generated by an AR(2) model with $a = 0.9$ and $b = -0.1$. Fitting $F^2(s)$ yields $g_1 = 0.16 \pm 0.14$ and $g_2 = 0.75 \pm 0.05$. Both parameters are g_2 correctly estimated with some uncertainty. CENTER: The process is generated with an unweighted superposition of two AR(1) processes with $g_1 = 0.2$ and for the second process $g_2 = 0.6$. Fitting the $F^2(s)$ with AR(2) yields $g_1 = -0.16 \pm 0.19$ and $g_2 = 0.50 \pm 0.10$ for this process. So the value of the stronger correlated model, g_2, seems to be captured by the fitted model. RIGHT: A deterministic sinusoidal signal. The period can be found correctly by fitting the fluctuation function.

the second order. The autoregressive process of order two, AR(2), introduced in section 2.2.3, is defined by

$$X(t) = aX(t-1) + bX(t-2) + \eta(t). \tag{5.1}$$

Just like in the previous chapter, $\eta(t)$ is any type of uncorrelated noise and a, $b \in \mathbb{R}$ are the AR parameters. The process is able to model relaxations and oscillations. The type of dynamics depends on the AR parameters. The autocorrelation function was given in section 2.2.3 as $C(t) = c_1 g_1^t + c_2 g_2^t$. Formally, it is a superposition of two AR(1) correlation functions. The theoretical fluctuation function $F^2_{a,b}$ for AR(2) can, therefore, also be expressed in terms of F^2_g of AR(1). A closer look at equation (4.10) and the AR(2) correlation function yields [F]

$$F^2_{a,b}(s) = \sigma^2 \Big(g_1 F^2_{g_1}(s) + g_2 F^2_{g_2}(s) \Big), \tag{5.2}$$

for the DFA fluctuation function of AR(2).

The AR(2) model describes the discrete version of a damped driven harmonic oscillator if the roots g_1, g_2 are complex numbers with positive real part. In the extreme case of $b = -1$, the correlation function is sinusoidal and does not decay. This correlation function describes a sinusoidal signal. Here the asymptotic scaling in DFA is not $\alpha = 0.5$ but $\alpha = 0$ in contrast to other choices of parameters.

The case of complex roots is highly relevant since oscillations are a fundamental feature of many real-world systems. Especially in atmospheric science, oscillations were found
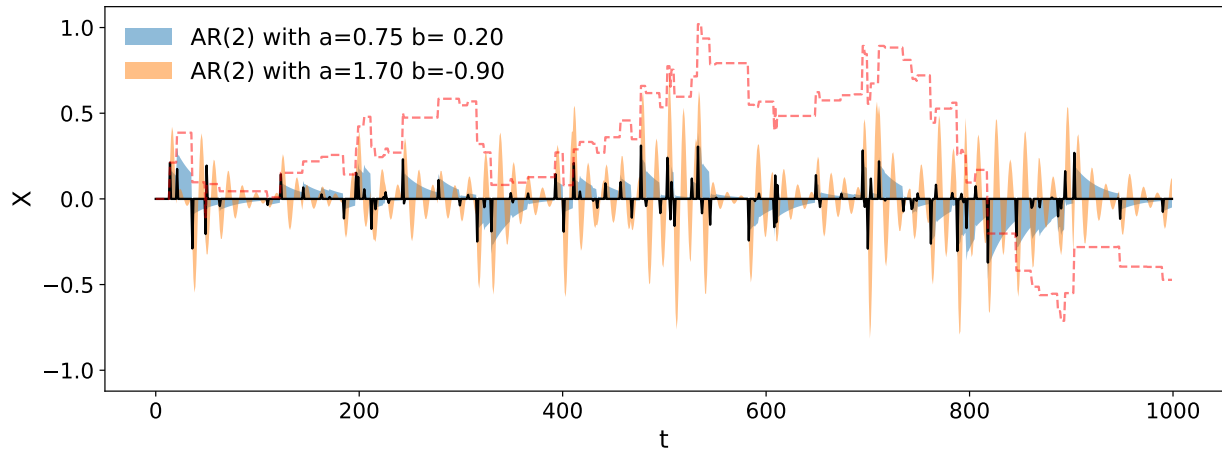
Figure 5.2

Two AR(2) processes with driven by the same realization of uncorrelated Poissonian noise η (black). One AR(2) process exhibits relaxations (blue; parameters: $a = 0.75$, $b = 0.2$) and one exhibits oscillatory dynamics (orange; parameters: $a = 1.7$, $b = -0.9$). The red dashed line marks dynamics driven by the noise without damping.

on numerous timescales. The AR(2) process (2.26) is the simplest way to model them.

Since the fluctuation function of AR(2) is given in equation (5.2), it can now be used for fitting, using equations (4.34) and (4.35). Thereby, periods or relaxation times can be inferred. The type of dynamics depends on the result of the fit. The minimization (4.34) has two free parameters a and b that are optimized simultaneously.

In figure 5.1, applications of the AR(2) fluctuation function fit to three example processes are shown. In the right panel, a sinusoidal signal is analyzed. The period can be reproduced by the fit with high accuracy. The parameter b is, in this case, -1. An estimation of the uncertainty is difficult in this case since the model is deterministic and no ensemble can be generated.

In the left panel in figure 5.1, application to a real AR(2) process with parameters $a = 0.9$ and $b = -0.123$ is shown. Alternatively, the model can be written as $(1 - 0.13B)(1 - 0.77B)X(t) = \epsilon(t)$ according to equation (2.28). The correlation function reads $C(t) = 1.08 \times 0.77^t - 0.08 \times 0.13^t$, so the dynamics exhibits no oscillations but only relaxations. Fitting yields the parameters $g_1 = 0.16 \pm 0.14$ and $g_2 = 0.75 \pm 0.05$. The errors are estimated by ensembles of 300 realizations of the process with 10000 time steps each. Accordingly, the method yields good fits of the parameters especially for the shorter characteristic time corresponding to g_1.

In the same way, the fluctuations of an unweighted superposition of two AR(1) processes, with $g = 0.2$ for the first process and $g = 0.6$ for the second process, are analyzed in the central panel of figure 5.1. The process can not be described by an AR(2) model, however, it can still be fitted by the AR(2) fluctuation function in order to get an approximate model. The obtained parameters are $g_1 = -0.16 \pm 0.19$ and $g_2 = 0.50 \pm 0.10$. The longer timescale, corresponding to g_2, is again close to the real value.
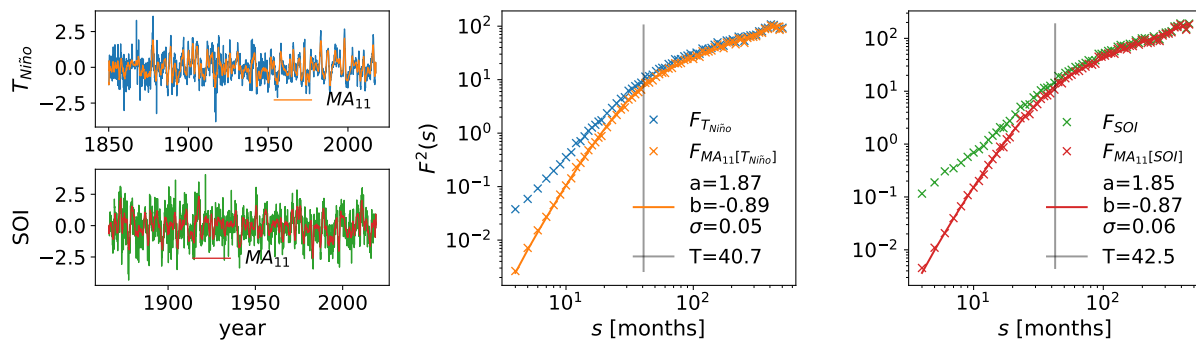
Figure 5.3

CENTER: Squared fluctuation function of monthly values of T_{Nino} (blue 'x') and 11 months averaged values of T_{Nino} (orange 'x'). The 11 month average is fitted by an AR(2) process. The fit (orange line) yields oscillatory behavior with period 40.7 months. Both time series are plotted in the TOP LEFT panel. RIGHT: Squared fluctuation function of monthly values of the SOI (green 'x') and 11 months averaged values of SOI (red 'x'). The 11 month average is fitted by an AR(2) process. Here the fit (red line) yields oscillatory behavior with period 42.5 months. Both time series are plotted in the BOTTOM LEFT panel.

In the last example, the interpretation of the result is not trivial. The real dynamics is not AR(2), however, the fit is close to the fluctuation function of the data. Therefore, it seems to reflect some properties of the investigated process. If AR(2) fits yield pure relaxations, it was proposed in [I] to report three alternative observables to g_1, g_2 and σ, which can easier be interpreted than the AR-parameters. These observables are the correlation time,

$$r = -\frac{h_1}{\log|g_1|} - \frac{h_2}{\log|g_2|}, \tag{5.3}$$

as well as σ_η of the noise and σ_X/σ_η. The latter is a measure of how strongly the AR(2) works as a low pass filter.

These parameters will become relevant in section 5.3. The interpretation of the AR(2) process as a control system, that leads to damping of perturbations is exemplified in figure 5.2. In comparison to the undamped dynamics, the AR(2) model brings the trajectories back to the mean value. The relaxation times depend on the parameters. For some parameters, when the roots g_1 and g_2 are complex, the process overshoots and becomes oscillatory. In the following sections, this parameter range is most relevant as it can model oscillations in atmospheric time series.

5.1.2 El Niño Southern Oscillation

The most prominent mode of climate variability is the El Niño southern oscillation [133] in the equatorial Pacific. The question of whether or not this really is an oscillatory mode, or rather a stable mode triggered by stochastic forcing, is still not finally solved
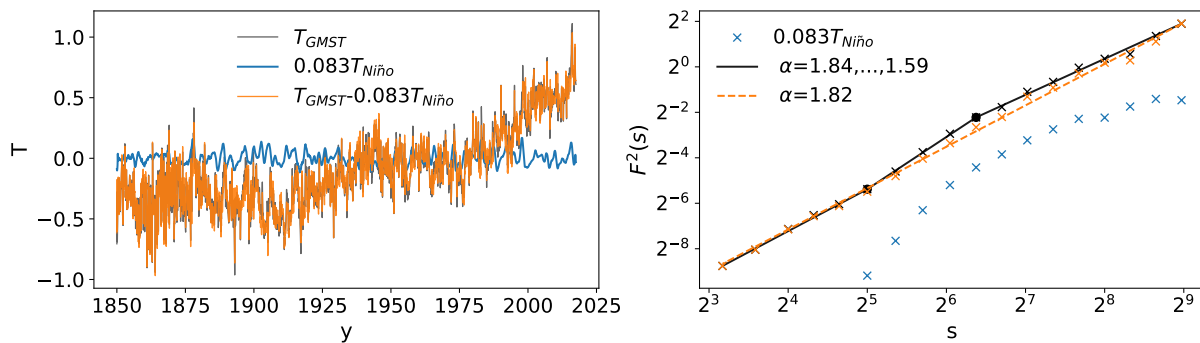
Figure 5.4

GMST time series (black), 11 month averaged ENSO time series with fitted pre-factor (blue) and the difference of both (orange). LEFT: data, RIGHT: fluctuation functions of all three time series.

[128]. Certainly, it is not a pure oscillation but noisy, which leads to the problem, that El Niño events do not appear regularly, but are, in fact, difficult to predict [75].

There are several measures of the El Niño southern oscillation. The traditional definition is the southern oscillation index (SOI). It describes the sea level pressure difference between Tahiti and Darwin, Australia. Instead of SOI, often the Niño 1–4 indexes are used, which are sea surface temperature averages over different regions in the equatorial Pacific [75]. Similar, but not the same, is the measure T_{Nino}, proposed by Thompson [124], which was used in section 4.2 as one of the main drivers of climate variability. It is further investigated here.

Due to strong short time variability, it is not plausible that the ENSO signal is equivalent to a damped oscillator driven by noise, i.e. an AR(2) model. In order to fit an oscillatory mode, the subseasonal variability has to be filtered out first. The easiest way to do this is by looking at running means. It turns out that an 11 months average works nicely for filtering the short time dynamics. Calculating 11 months means yields a steeper slope of the fluctuation function for short s, while longer timescales remain unaffected. The DFA fluctuation function is fitted in figure 5.3 in the central panel. A similar plot was published in [F]. The goodness of the fit suggests that it is indeed only dominated by one oscillatory mode for this timescale. The AR(2) fit yields a noisy oscillation with a period of 40.7 months, which is a bit more than three years and a bit less than the low frequency El Niño period calculated in [44] with a different method.

An identical analysis can be applied to the SOI time series. Here fitting the fluctuation function of the 11 month averages yields an AR(2) model with a period of 42.5 months, as shown in figure 5.3 in the right panel. The result is close to the one obtained from T_{Nino}, so apparently, both approaches measure the same phenomenon. The deviation of 1.8 months gives an idea of the uncertainty that comes with the method of fitting fluctuation functions. The data for both time series, SOI and T_{Nino}, as well as the 11 month running means, are plotted on the left side of the figure.

ENSO also affects the fluctuation function $F(s)$ of temperature time series by causing a concave crossover [91]. It is similar, yet less pronounced, than the seasonal cycle. The effect on global mean surface temperature anomalies is demonstrated in figure 5.4 following [F]. In the left panel the time series of GMST anomalies, as well as the 11 month averages of T_{Nino}, are shown in a monthly resolution. In the right panel, the fluctuation functions are shown. Around the ENSO period of 3-4 years, there is a crossover in GMST. Now similar to the approach in section 4.2, the 11 months averages of T_{Nino} are fitted to the GMST time series. The fit is subtracted with the obtained pre-factor. The resulting time series looks very similar to the original GMST time series, as shown in the left panel of figure 5.4. However, in the fluctuation function (right panel) the effect is visible. The fluctuation function of the signal, where T_{Nino} was subtracted does not show a clear crossover anymore. It is significantly less curved than the original GMST fluctuation function. Hence, El Niño detrending here works similar to removing the seasonal cycle. If the fit was perfect and ENSO was really independent of other features in the GMST time series, the fluctuation functions $F_{\tilde{T}}^2$, $F_{T_{Nino}}^2$ and $F_{\tilde{T}-T_{Nino}}^2$ would follow the superposition principle (4.18).

According to the presented analysis, El Niño events are not purely driven by noise, but there is an underlying dynamics and a characteristic period. This raises hope that the phenomenon is predictable, which would of huge benefit for the affected regions. In fact, there are some promising results already on long term forecasts of El Niño events [26, 75].

5.1.3 The superposition principle and fitting DFA

The previously introduced example of ENSO as one component of GMST raises the question if a time series with multiple characteristic timescales can also directly be analyzed by fitting fluctuation functions. As the examples of AR(1) and AR(2) processes illustrate, DFA highlights characteristic timescales as local maximum values of the negative curvatures of the fluctuation function. The easiest assumption of how such a system with multiple characteristic timescale might be constituted is a superposition of processes

$$v(t) \approx \sum_i X^{(i)}(t), \tag{5.4}$$

with one characteristic timescale each. Note that such a model is usually a simplification that only approximates the real dynamics. However, for example, in the case of atmospheric time series, the underlying dynamics seems to be well approximated by a superposition due to a separation of timescales. Examples will be discussed in the following sections. It was already shown in section 4.1.4 that for processes of type (5.4) the DFA fluctuation function follows the superposition principle (4.18)

$$F_{\Sigma_i X^{(i)}}^2(s) = \sum_i F_{X^{(i)}}^2(s). \tag{5.5}$$

The processes $X^{(i)}(t)$ are simple models with one characteristic relaxation time or oscillatory mode. These timescales τ_i are assumed to be ordered in time according to

$$\tau_i < \tau_j \quad \text{for} \quad i < j. \tag{5.6}$$

Examples are the already discussed autoregressive processes of order one or two. There are two conditions for the method to work. Firstly, the $X^{(i)}(t)$ have to be pure oscillations or short range correlated, i.e. low pass filters. Secondly, the characteristic timescales of the processes $X^{(i)}(t)$ have to be sufficiently well separated. Then up to s_{τ_i}, corresponding to the characteristic timescale τ_i of process $X^{(i)}(t)$, the contributions of all processes $X^{(j>i)}(t)$ are negligible

$$F_{X^{(i)}}^2(s) \gg \sum_{j>i} F_{X^{(j)}}^2(s) \quad \forall s \leq s_{\tau_i}. \tag{5.7}$$

First, the full DFA fluctuation function $F_v^2(s)$ is calculated with approximately logarithmic steps of s. These are generated by starting from $s_{\max}$, smaller or equal to one fourth of the length of the time series, and following the recursion $s_i = \max_{s \in \mathbb{N}} s \leq s_{i+1}$. Then, the crucial part of the method is the choice of the model that should be fitted. This can be done by visual inspection of the fluctuation function or trial and error. The analysis is then carried out by starting on the shortest time scale $X^{(1)}(t)$ and then repeating the three following steps until the fluctuation function is completely described in the available time range:

- $F_{X^{(i)}}^2(s)$ is fitted for short s with $s_{\min} \leq s \leq s_{\tau_i}$ to the remaining fluctuation function $F_{v-\Sigma_{j<i}X^{(j)}}^2(s)$ (in the first step it is the full fluctuation function $F_v^2(s)$). The time range and the model should be chosen such that the model has as few as possible parameters and the considered range of s is as long as possible without significant deviations in F of the data and the model in the considered range. Then

- subtract the fitted fluctuation function from the original fluctuation function for all s

$$F_{v-\Sigma_{j\leq i}X^{(j)}}^2(s) = F_{v-\Sigma_{j<i}X^{(j)}}^2(s) - F_{X^{(i)}}^2(s). \tag{5.8}$$

For small s, the values of $F_{v-\Sigma_{j\leq i}X^{(j)}}^2(s)$ should all be zero. In practice, some will be slightly above zero and some will be slightly smaller than zero. Since negative values are not allowed in a log-log plot,

- choose a new time range by dividing s and $F(s)$ by some parameter k. This corresponds to downsampling of the data, however, it is only done on the level of the fluctuation function.

$$s_1 \to s_k k, \qquad F \to F/k, \qquad \text{i.e. } F_{v-\Sigma_{j\leq i}X^{(j)}} \to F_{\lfloor v-\Sigma_{j\leq i}X^{(j)}\rfloor_k} \tag{5.9}$$

such that the rescaled fluctuation function has positive values for all $s > 3$, i.e. in the range where the next analysis is performed.

This method was first published in [G]. In the following section, the analysis is applied to a time series where the previously mentioned El Niño Southern Oscillation is present.
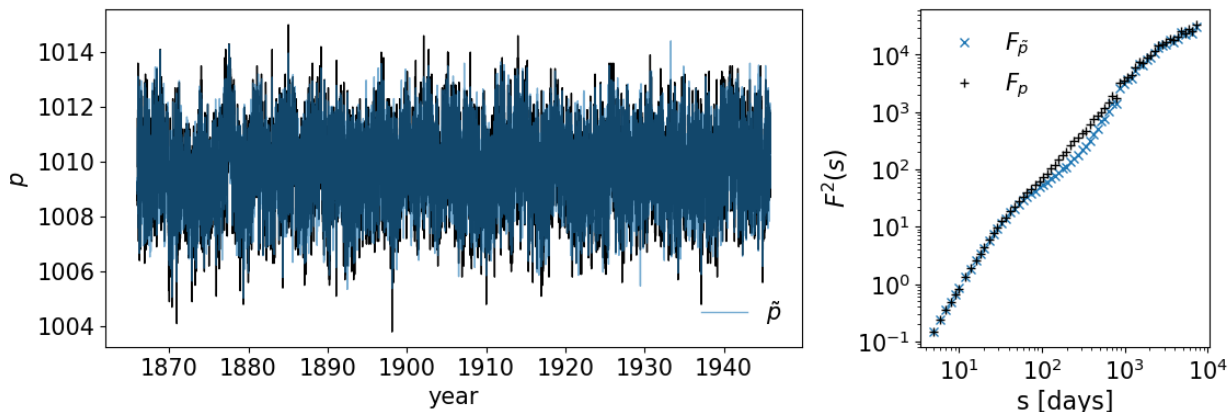
Figure 5.5

LEFT: Sea level pressure from Kemayoran. The black lines mark the recordings while the anomalies are plotted as colored dots. RIGHT: Squared fluctuation functions of both time series.

5.1.4 Oscillations in Pacific sea level pressure

The recordings of sea level pressure from Kemayoran, Indonesia, are a good example of a time series that exhibits fluctuations on more than one characteristic timescale. They were already presented in [H]. The station is located in the region of the El Niño southern oscillation. Therefore unlike some European pressure time series, it can not be described just by one AR(1) process. In fact, climate variability in the equatorial Pacific is even more complex than the ENSO discussion in section 5.1.2 suggest. On timescales smaller than one year even the SOI itself does contain a second characteristic oscillation [131]. This is the Madden-Julian Oscillation [76], which is also typical for the equatorial area and has a period around 40-50 days.

The data in Kemayoran, Indonesia was recorded between 1.1.1866 and 16.11.1945 with only few missing values. The time series and the complete fluctuation function are shown in figure 5.5. Since the data exhibits some seasonality and the ENSO frequency is close to one year, it is necessary to analyze anomalies $\tilde{p}(t)$ instead of $p(t)$. They are calculated as described in section 2.2.2.

Now the method, suggested in section 5.1.3, shall be applied. Figure 5.6 shows the analysis in three steps. First, in the left panel, an AR(1) process $X(t_d)$ with one-day-time-steps is fitted to the short s behavior of the fluctuation function $F_{\tilde{p}}(s)$. This implies exponential decay of correlations for short times. It is usually a good description of the short time behavior of temperature or pressure time series (see e.g. sections 4.3.4 and 5.2.2). The fit is a bit difficult here because the Madden-Julian oscillation period is quite short compared to the relaxation time and, therefore, the condition of sufficiently well separated characteristic timescales is almost not fulfilled. However, the fit works, if it is performed in the small window $s < s_{\max} = 10$, highlighted in the figure in gray. It yields a relaxation time of 3.63 days. The fluctuation function F_X^2 can be subtracted from $F_{\tilde{p}}^2$.
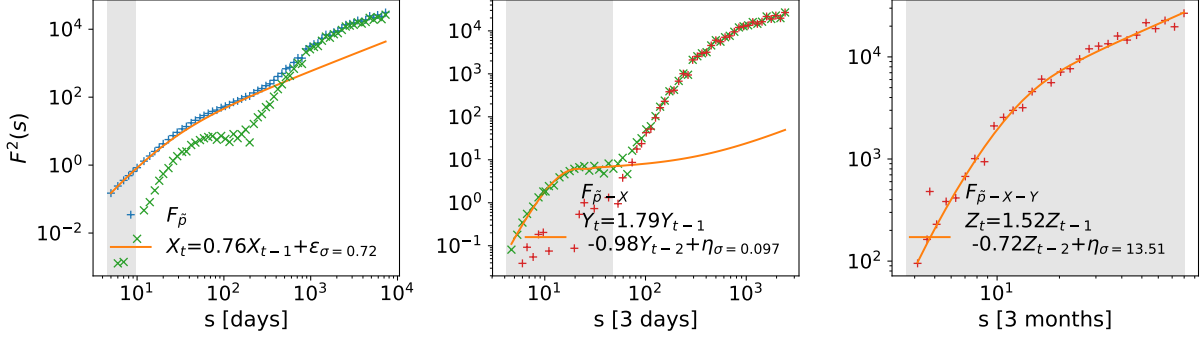
Figure 5.6

LEFT: Full fluctuation function $F_{\tilde{p}}$ from Kemayoran (blue +), fitted by AR(1) model F_X (orange line) in the range highlighted in gray and the result for $F_{\tilde{p}-X}$ (green x). CENTER: $F_{\tilde{p}-X}$ fitted by F_Y (orange line) in the gray area and difference $F_{\tilde{p}-X-Y}$ (red +) RIGHT: $F_{\tilde{p}-X-Y}$ (red +) and fitted AR(2) fluctuation function F_Z (orange line).

For the second step, a suitable rescaling of the fluctuation function is needed, such that the part of the fluctuation function is neglected that was well described by F_X^2 and therefore is close to zero in $F_{\tilde{p}-X}^2(s_d)$. A good time step duration for the second fit seems to be three days. So a 3-days timescale is created by multiplying $s_{3d} = 3s_d$ and dividing $F_{\tilde{p}-X}$ by 3. Now points $s < 4$ can be neglected. The remaining fluctuation function $F_{\tilde{p}-X}^2(s_{3d})$ is shown in the central panel of figure 5.6. It is fitted by an AR(2) process, that can model the Madden-Julian oscillation. Here it is not difficult to find a suitable s-range for the fit since the oscillation is clearly visible by the concave shape of the fluctuation function. The fit yields an oscillation with a period of 42.8 days and only small noise. Again, the fit is subtracted and thus the remaining fluctuation function is $F_{\tilde{p}-X-Y}^2(s_{3d})$.

The third fit in the right panel of figure 5.6 is performed with a model with time steps of 3 months s_{3m}. Therefore, s is multiplied by 30.417 and $F_{\tilde{p}-X-Y}^2$ is divided by the same number. The resulting $F_{\tilde{p}-X-Y}^2(s_{3m})$ is then fitted by an AR(2) process that models ENSO. The fit describes the complete remaining fluctuation function well, thus no additional slow process has to be considered. It yields noisy oscillations with a period of 40.7 months, which is very much compatible with the results from section 5.1.2.

So the fluctuations of $\tilde{p}(t)$ can be described by a superposition of three processes

$$\tilde{p}(t) \approx X(t_s) + Y(t_{3d}) + Z(t_{3m}). \tag{5.10}$$

Here $X(t_s)$ is an AR(1) process, that describes the short correlations, $Y(t_{3d})$ is an AR(2) process that accounts for the Madden-Julian oscillations, and $Z(t_{3m})$ is also an AR(2) process modeling ENSO. It is a simple model that captures the complete power of the fluctuations and can be used as an approximation of the dynamics with qualitatively correct characteristic timescales.

5.2 Decomposing European temperature variability

In section 1.3 the literature on climate variability was briefly summarized. The main observation is that there are two perspectives on this topic. Parts of the literature concentrate on oscillatory modes and others on the apparent scaling of fluctuations, indicating long range correlations.

In the following, the fluctuations of temperature measurements for European stations will be analyzed from both perspectives with DFA. In this way, the results that researchers obtain for different methods can be brought together. A possible interpretation is that both, the scaling and the oscillations on multiple timescales, are two sides of the same coin. Obtaining the correct fluctuation functions with both a long range correlated model and a short range correlated model with several characteristic timescales can explain the 'missing quadrillion' in variance mentioned by Lovejoy [73]. The analysis with a long range correlated model is presented in section 5.2.1. The short range correlated model is introduced in section 5.2.2.

There is a large number of stations with several decades of daily recordings of temperature. Station data of temperature and pressure in Europe [120] can be downloaded from `www.ecad.com`. There also is a long history of investigations on both individual stations or grid data. One important result is the finding of a 7-8 year cycle for various stations and other records [99, 95, 115] and its spacial patterns [50, 98, 62]. The origin of this feature is not known.

Here an investigations on a large number of stations will be presented in sections 5.2.3 and 5.2.4. The analysis can be performed in an algorithmic way. The results of this part are published in [G]. The figures 5.8, 5.9, 5.10, 5.11, and 5.12 are taken from the publication. Map backgrounds were made with Natural Earth (`naturalearthdata. com`).

5.2.1 The ARFIMA model for temperatures

In [84] it was shown by Massah and Kantz, that an ARFIMA(1, d, 0) model works well for describing temperature anomalies. In this way, the measured scaling exponents can be modeled, as well as some stronger short range correlations that people experience (the weather of tomorrow is probably similar to the weather today). The parameter d, that describes the long range correlations, is typically quantified around 0.65 [39]. It should be pointed out that it is essential to use temperature anomalies $\tilde{T}$ for this kind of scaling analysis, where the seasonal cycle is subtracted from the data. Otherwise, the yearly cycle is dominant in the data and also in DFA [33].

In figure 5.7, temperature data from Potsdam is analyzed. In the left panel, the time series of T and of the anomalies $\tilde{T}$ is plotted. In the central panel, the fluctuation function $F_{\tilde{T}}^2(s)$ is shown. It is analyzed with the traditional DFA, i.e. a power law is fitted that describes the asymptotic behavior. The exponent is $2\alpha = 1.3$ for the squared fluctuation function. So α is in fact 0.65 as found in previous investigations [84].

Now it can be verified that if the fluctuation function is really compatible with ARFIMA(1, 0.15, 0). In order to do so, in the right panel of figure 5.7, the squared
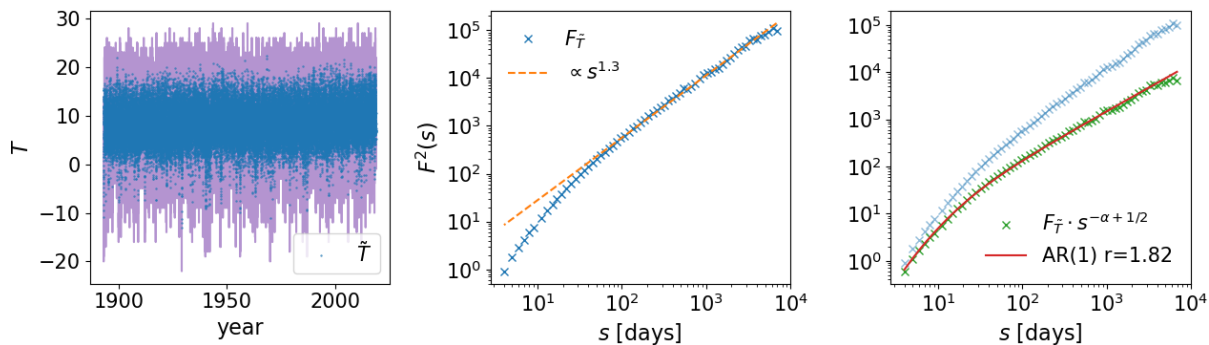
Figure 5.7

LEFT: Temperature data from Potsdam station. The measured values are displayed with solid lines in the background. The anomalies are the dots in the foreground. CENTER: fluctuation function of the temperature anomalies (blue 'x'). The fit (dashed line) indicates asymptotic scaling with $\alpha = 0.65$. RIGHT: fluctuation function of the temperature anomalies (blue) and fluctuations of the rescaled values $F_{\tilde{T}} \cdot s^{-\alpha + 0.5}$ (green 'x'). The rescaled values are fitted with the fluctuation function of an AR(1) model (red line). The function is well fitted. The analysis yields a relaxation time of $r = 1.82$ days.

fluctuation function is multiplied by the factor $s^{-2\alpha+1}$. This follows the idea of understanding combinations of processes in section 4.1.4 and equation (4.23). ARFIMA(1, d, 0) is such a combination of an ARFIMA(0, d, 0) process that is put into an AR(1) process. So the fluctuation function $F^2 s^{-2\alpha+1}$, where the scaling d was removed, should now be the fluctuation function of an AR(1) process. The theoretical AR(1) result is fitted in the right panel of the figure and indeed matches very well.

It can be concluded that the fluctuations of the Potsdam temperature anomalies in the available time range are well described by an ARFIMA(1, 0.15, 0) model with AR parameter $g = 0.58$. The correlation time of the AR(1) model is very short (1.82 days) because the memory is partly due to the ARFIMA parameter d.

5.2.2 The superposition model for temperatures

The ARFIMA model, applied in the previous section, is one approach that explains the temperature fluctuation functions and features long range correlations. In this section, an alternative approach is shown that explains the fluctuation function with a superposition of simple short range correlated processes. The idea is the same as in section 5.1.4.

Instead of looking at the anomalies, it is now beneficial to look at the real data T. Here, in contrast to the analysis in the previous section, the values for the 29th of February in leap years do not have to be removed. Since the goal of the method is to identify characteristic timescales, the seasonal cycle is expected to be one of the main features. The hope is by looking at T instead of $\tilde{T}$, the effect of the seasonality on the fluctuations can be quantified more accurately than by calculating the mean values for
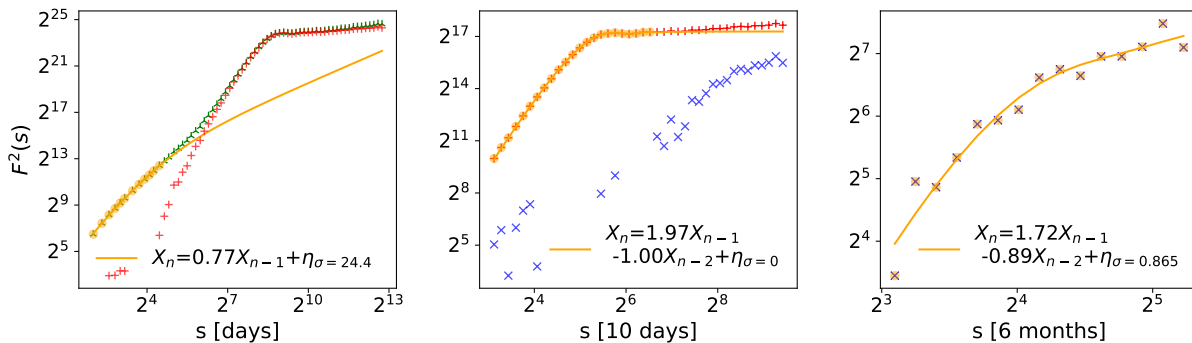
Figure 5.8

Decomposition of the fluctuation function of Potsdam temperatures. The orange curves represent fits to the fluctuation functions. The orange circles mark the points for which the fit is performed. LEFT: short time behavior F_X^2 fitted to the full fluctuation function F_T (green '$\curlywedge$'), which yields a relaxation time of 3.8 days. The red '+' denotes $F_{T-X}^2 = F_T^2 - F_X^2$; CENTER: seasonal cycle F_Y^2, fitted to F_{T-X}^2. The resulting F_{T-X-Y}^2 is plotted as blue 'x'; RIGHT: interseasonal variability F_Z fitted to F_{T-X-Y}.

each calendar day. This is because the seasonal cycle is in reality not static, but the phase oscillates [32]. The beginning of the vegetation period, for example, is not on the same date every year.

The fundamental assumption of this section is that the time series can be written as a superposition of three processes. The timescales are picked at 1 day, 10 days and 6 months. These choices are made by trial and error and the exact values are not essential. The simplified model for temperature reads

$$T(t) \approx X(t_d) + Y(t_{10d}) + Z(t_{6m}), \tag{5.11}$$

where $T(t_{10d}) = \lfloor T(t_d) \rfloor_{10}$. All three constituents should be described by one of the introduced autoregressive models. It should be stressed that the independence of the three processes in this simple model does not mean that there are really no interactions between them. In fact, such interactions between processes on different timescales are a current subject of research [95]. However, the effect of the interactions on the fluctuations can be considered negligible compared to the total power of the fluctuations. This is justified mainly by the success of the decomposition.

In figure 5.8, the complete decomposition is shown for measured temperature data from Potsdam. As it was already discussed for the ARFIMA analysis, temperature correlations for short times decay exponentially. The behavior is apparent in the left panel of figure 5.8 in the full fluctuation function $F_T(s)$ for small s. So the fluctuations can be well described by an AR(1) model $X(t_d) = gX(t_d - 1) + \eta(t_d)$. The noise term $\eta(t_d)$ seams to be close to Gaussian noise as are the temperatures [84]. They are not further analyzed here. The variance σ_η^2 is a free parameter. For simplicity, one can look at the standard deviation σ_X of the process instead of σ_η. They are related by equation (2.25).

For Potsdam temperatures the fit of the AR(1) model to F_T for small s yields a relaxation time of 3.8 days (or $g = 0.77$). Other subseasonal timescales are not pronounced in the fluctuation function. This lack of characteristic timescales is known. It is the reason why predictions on this timescale are difficult [45]. Now the fitted AR(1) fluctuation function F_X^2 can be subtracted from F_T^2.

Now consider time steps of 10 days by dividing F_{T-X} and s_d by 10 and removing all points with s_{10d} smaller than 3. It is easily visible in the central panel of figure 5.8 that the smallest remaining characteristic timescale is the seasonal cycle, which is the dominant feature of the signal. Although it is not perfectly periodic, it has a phase-locked frequency of $1/365.25$ days, which is enforced by the periodic driving of the sun. So the effect of the yearly cycle is approximated by the fluctuation function of an AR(2) process with $b = -1$, which is equivalent to a sinusoidal signal

$$Y(t_{10d}) = \frac{\sigma_Y}{2} \sin\left(\frac{2\pi t_{10d}}{365.25d} \right). \tag{5.12}$$

It does not describe the phase fluctuations, but accounts for the full power power of the seasonal cycle as it is fitted to the fluctuation function. In the central panel in figure 5.8, one can see that the fit works very well for the fluctuation function. The remaining fluctuation function F_{T-X-Y}^2 is calculated by again subtracting the squared fluctuation function.

The timescale is now changed to half a year (182.625 days) by dividing both s_{10d} and F by 18.2625 and removing points with $s_{6m} < 3$. The remaining fluctuation function is shown in figure 5.8 in the right panel. The question is now: which type of dynamics can describe the interannual variability up to a bit more than a decade (the time range available)? There is one interesting feature that was found for some European stations, namely a 7-8 year oscillation. It was first detected in the record of central England temperatures [99] and since then also in several other time series of European temperatures [62].

Such a frequency mode could be captured by an AR(2) model $Z(t_{6m}) = aZ(t_{6m}-1) + bZ(t_{6m}-2) + \eta(t_{6m})$. For Potsdam, the fit seems to show good agreement with F_{T-X-Y}^2. Naturally, uncertainty in DFA grows for large s since the averaging is done over fewer windows. The fit yields a period of 7.5 years. The full fluctuation function is explained by the presented approach.

In addition to the analysis for Potsdam, temperature fluctuations are decomposed for three more stations in figure 5.9. The stations are Arkhangelsk, Bologna, and Oslo Blindern. They were picked because their results clearly differ from each other. Oslo Blindern is another example for a station where an oscillation could be identified with a measured value of $\tau = 8.2$ years, very similar to Potsdam. The other two stations do not exhibit a significant frequency mode in the interannual variability. For Arkhangelsk, the algorithm yields AR(1)-like interseasonal dynamics with a relaxation time of approximately 2.5 years. For Bologna, it yields a slow oscillation, with period length 39 years, due to the steep increase of the fluctuation function. Since 39 years are outside the considered time range, this number is not a reliable fit. The interpretation of such a
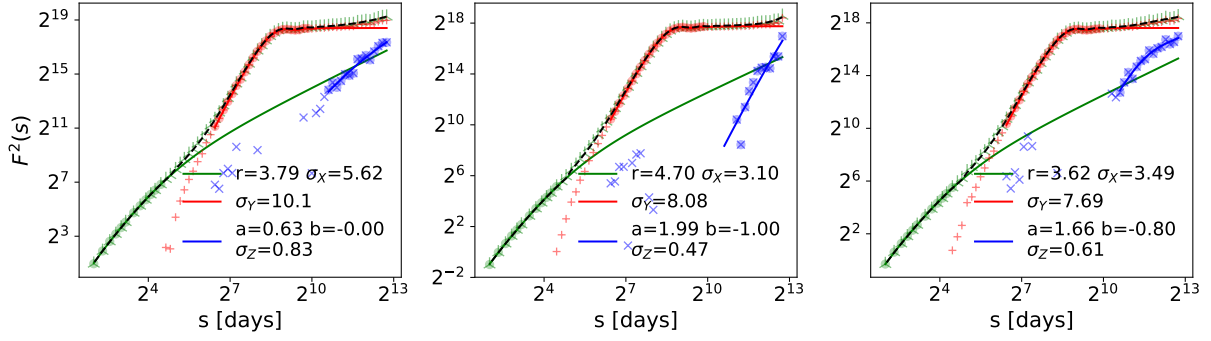
Figure 5.9

Decomposition of F_T for different stations. The colorcoding is identical to the previous figure (F_T^2: green '⅄'; F_{T-X}^2: red '+'; F_{T-X-Y}^2: blue 'x'), however, here the full decomposition (all three steps) is shown in one panel. The fits are drawn in the same color as the data they a fitted to. In order to increase clarity, the fluctuation functions are not rescaled after each step. The dashed line in each panel represents the result F_{X+Y+Z}^2 of the decomposition. LEFT: Arkhangelsk, where F_Z^2 is best described by an AR(1) model; CENTER: Bologna, where F_{T-X-Y}^2 is very steep, which means that the next oscillatory mode is outside the presented s-range; RIGHT: Oslo Blindern, where an oscillatory mode with period length $\tau = 8.2$ years was detected.

behavior is that there is some slow process with an unknown characteristic time beyond the maximum of the considered range (around 15 years).

5.2.3 The short time behavior and sea level pressure

Using the model presented in section 5.2.2, the temperature variability for different European stations can be compared. In the first step, the data is fitted algorithmically. Six parameters, r, σ_X (for the first fit), σ_Y (for the second fit), a, b, and σ_Z (for the third fit), are obtained for each station. Only datasets are considered, which cover at least 60 years of (non-blended) daily temperature measurements and have at most 6.7% of missing days. In an area around Germany with very good coverage, the threshold is set a bit higher at 75 years of recordings. In total, 336 stations were considered. The results of the fit for the parameters r, σ_X, σ_Y, and σ_Z are presented in this section. The results for the interannual variability are discussed in the following section.

The short time variability is modeled by an AR(1) process that is fitted to the short s behavior. In section 4.3.4, it was shown that the sea level pressure for European station behaves very similar in this regard. Moreover, the sea level pressure fluctuation functions are, in many cases, already completely described by this model. Now one might ask if the relaxation times for temperature and pressure at the same station are equal. The question can be answered by fitting the model, $X(t_d)$, not only to the temperature time series but also to the corresponding sea level pressure time series.

The results are shown in figure 5.10. Not all stations with long temperature records
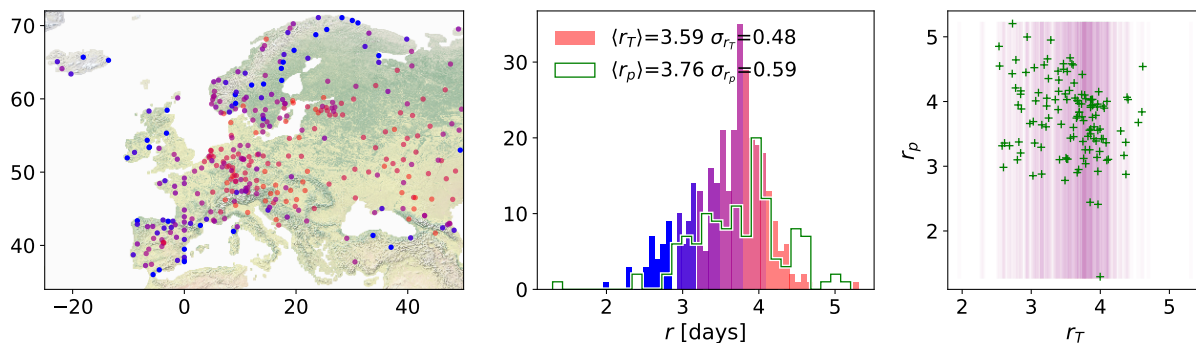
Figure 5.10

LEFT: Color plot of temperature relaxation times r_T, obtained from the AR(1) parameter (2.21) for each station. The color map can be found in the histogram in the CENTER, where it is compared with the histogram of the results of the same procedure for pressure data r_p (green line). The distributions of r_T and r_p are similar to each other; RIGHT: r_T and r_p are plotted against each other (green '+') for stations where pressure data is available. The plot shows, that there is no visible correlation between both. The purple heat map in the background represents values of r_T for all stations, i.e. also stations with no pressure recordings available.

also measure sea level pressure. If there are at least 4 years of pressure recorded at the station, the fit is performed for both time series. Otherwise, it is done for temperature only. The fluctuation function is calculated for a set S of s-values, constructed from the maximal value $s_{max} = 365 \cdot 75$ and the recursion $s_i = \max_{s \in \mathbb{N}} s \leq 0.9 s_{i+1}$, which yields a logarithmic scale. For datasets with less than $4 \cdot s_{max}$ recorded days the values of s which are larger than 1/4th of the length of the time series are ignored. For the analysis, the pressure and temperature fluctuation functions are fitted for $s \leq 25$ days.

The results for the relaxation time r of temperatures exhibit clear regional patterns, visible in the left panel of figure 5.10. The correlation time is short in the west and the north of Europe close to the Atlantic. It is longer in central and eastern Europe and around the Baltic sea. The histogram of relaxation times is shown in the central panel. The average relaxation time is $r_T = 3.59$ days. The standard deviation of 0.48 is mostly due to regional differences and not a measure of the uncertainty of the method.

For pressure data, the histogram of obtained values (figure 5.10 center) looks very similar to the one obtained for temperatures. The obtained values are around $r_p = 3.76 \pm 0.59$. However, regional patterns differ. Pressure values relax faster in the south compared to the north of Europe. The right panel of the figure shows that there is no visible connection between r_T and r_p for the individual stations. So even though both describe short time weather patterns, the short time dynamics of pressure and temperature is locally not equivalent.

The variance of the noise $\eta(t_d)$ in the short time dynamics is low close to the coasts and has its highest values in Russia and Scandinavia in more continental climate. The regional patterns are very strong here (see upper left panel of figure 5.11 and lower left
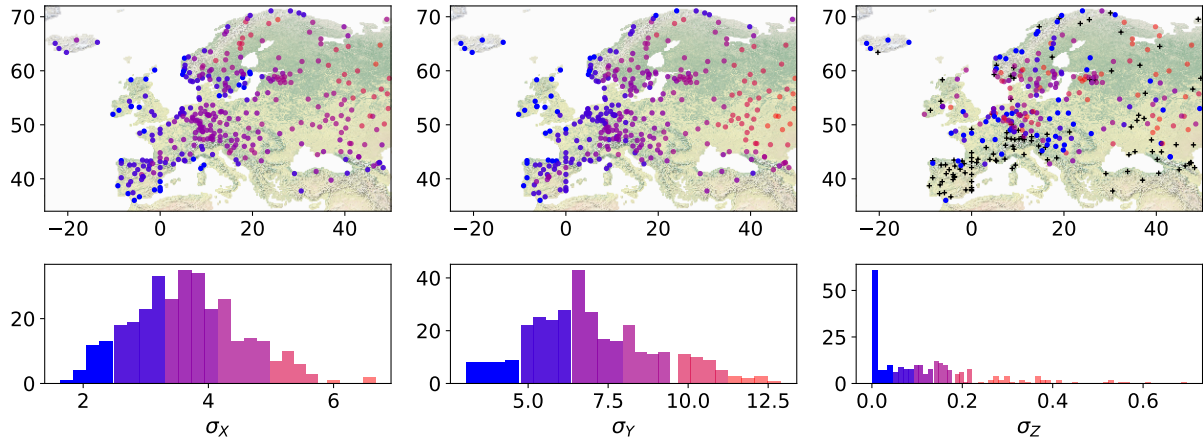
Figure 5.11

The standard deviation in all three processes X, Y and Z represented by the parameters σ_X (LEFT), σ_Y (CENTER), and σ_Z (RIGHT). For σ_Z stations where no oscillatory behavior in the range of 2.5 to 30 years was found are neglected and marked by black crosses. TOP: Results for all stations. BOTTOM: Histograms with identical color-coding.

panel for histogram). The seasonal cycle, i.e. the process $Y(t_{10d})$, is fitted for $8 < s < 38$ times 10 days. The color plot in the central panels of figure 5.11 looks almost identical to the plot of the short time variance. The seasonal cycle is most pronounced far away from the coasts.

The AR(2) model $Z(t_{6m})$ is fitted for $s > 8$ [6 months] ($= 48$ months). The variances of the process $Z(t_{6m})$ is shown in the right panels of the figure. Here the result depends on the obtained type of dynamics, i.e. oscillatory dynamics leads to a larger variance as pure relaxations. In order to have comparable results, only values for processes with an oscillation period of 5 to 11 years are shown in colors. Regional patterns are recognizable, however, they are more complicated than for the other processes.

5.2.4 The 7–8 year cycle

When fitting F_Z of the AR(2) model to F_{T-X-Y}, the dynamics might be very different from each other depending on the station. The only interpretable results are oscillatory modes within the considered parameter range because the fit works best if it sees the crossover in the fluctuation function. So the analysis focuses on detecting such features. Consequently, two cases are distinguished. The set Ω_+ shall contain all stations where an oscillatory mode with a period τ between 10 and 22 half years is detected. The set Ω_- contains all other stations.

There might be three reasons for a station to be classified in Ω_-, i.e. three types of dynamics other than the oscillatory behavior. This is the case if

- the slope is $\propto s^{0.5}$ or even flatter and the algorithm indicates an oscillation with a very short period,
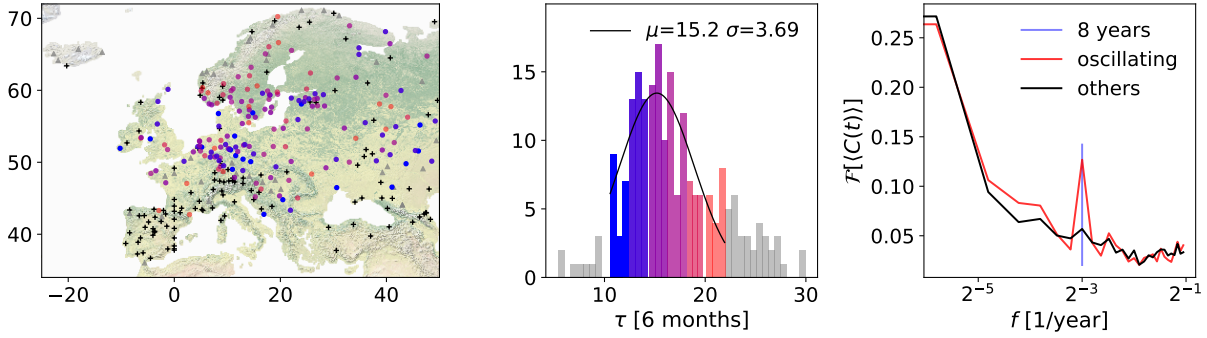
Figure 5.12
LEFT: Results of period times τ for all stations. Fitting parameters that do not yield a period or a period far off the investigated timescale are marked as black crosses. Periods between 10 and 22 times 6 months are marked with colored dots. Other periods between 5 and 30 times 6 months are marked in gray. CENTER: Histogram of colored and gray points. A Gaussian curve is fitted to the colored part via the method of least squares, yielding 7.6 ± 1.8 years as the average period time. RIGHT: Power spectrum for stations in Ω_+ with 5 to 11 year period (red) vs power spectrum for all other stations in Ω_- (black).

- the slope is very steep ($\propto s^{1.5}$) and the algorithm indicates an oscillation with a very long period, or

- the station lies in between, where the AR(2) model has real roots indicating no oscillation.

An example for the second type is Bologna, shown in figure 5.9 (center). An example of the third type is Arkhangelsk analyzed in the left panel of the figure. All three cases have in common that there is no sharp crossover in the fitted s-range and therefore statements about periods can only be made with high uncertainty.

Out of the 336 stations that were considered in total, 171 belong to Ω_+, i.e. exhibit the interseasonal oscillation. So in 51% of the analyzed stations, an oscillation is observed, see figure 5.12. The values are shown in the histogram in the central panel. The mean value is at $15.2 \cdot 6$ months, which corresponds nicely to the previously found 7-8 year cycle. The result for many stations is close to this mean value. The fit of a Gaussian curve to the histogram was done by the method of least squares and not by calculating the first and second moment. This is because, otherwise, the result would strongly depend on the choice of the considered range (5 - 11 years) for the classification as Ω_+.

The categorization into the two subgroups Ω_+ and Ω_- uncovers regional patterns. The mode was found by the performed analysis in Britain, the south of Scandinavia and central Europe from the North Sea to the Alps and Poland. It is also visible in parts of eastern Europe. In both ways there are some exceptions, i.e members of Ω_+ or Ω_- that are surrounded by the other group. These exceptions are likely to be false classifications since the error of the method seems to be quite high. It can be estimated

by looking at the histogram in the central panel of figure 5.12. Since the values for τ do not show clear regional patterns, the standard deviation of the values is a measure of the uncertainty of the method for the period τ. The hypothesis is that the real dynamics only has one characteristic timescale which is measured for several stations with some uncertainty. This uncertainty is determined to be 1.8 years which is more than spectral methods typically have [12].

So the algorithm classifies some stations as Ω_+, where the 7-8 year period is significant and some stations are classified as Ω_-, where it is not significant. The question is now whether these classifications are due to low sensitivity or large errors of the method, or if there are indeed two or more significantly different types of dynamical behavior in European temperatures. In order to answer this question, the conditionally averaged power spectra,

$$P(\nu) = \mathcal{F}[\langle C(t) \rangle_\Omega], \tag{5.13}$$

are calculated for both classes. Here $\mathcal{F}$ is the Fourier transform and the brackets $\langle \rangle$ denote the average over all stations in the set Ω_+ or the average over all stations in the set Ω_-. This validation can be seen in figure 5.12 in the right panel. It shows clearly that stations Ω_+ exhibit a peak at 8 years in the power spectrum while the others (Ω_-) do not. So the method of fitting fluctuation functions is able to classify the dynamics in a meaningful way.

Even though accuracy is lower than for some spectral methods the proposed method of fitting fluctuation functions has some advantages. It provides a full data model of the time series that account for the full power of oscillations. It can also be applied in an algorithmic way. It can deal with different types of dynamics like relaxations oscillations and noisy oscillations. In the following sections, the power of the method will be further tested by applying it to a completely different kind of system.

5.3 Power grid frequency fluctuations

All applications presented so far are investigations on atmospheric time series. Inferring characteristic time scales and generating data models using DFA fluctuation functions seems to work well for sea level pressure and temperature time series from various locations. In these final sections, the method is applied to a completely different problem. Here measured power grid frequency data is analyzed in the same way as the temperature measurements before. In contrast to climate data, the information that these fluctuations contain does not give insights into the physics of the atmosphere, but they reflect human decisions, habits and the construction of the human-made power grids.

Nevertheless, studying the power grid frequency fluctuations is highly relevant. It can give insights into how stable the grid is [72]. European power grids use an operational frequency of 50 Hz. If the grid frequency deviates too much from this value the system can become unstable and power outages might occur. Studying the statistical properties of the grid dynamics can help to understand the likelihood and causes of such events.

Therefore, stochastic time series models were already proposed [46], in addition to network models (e.g. [10, 80]). Both implement some known features and principles
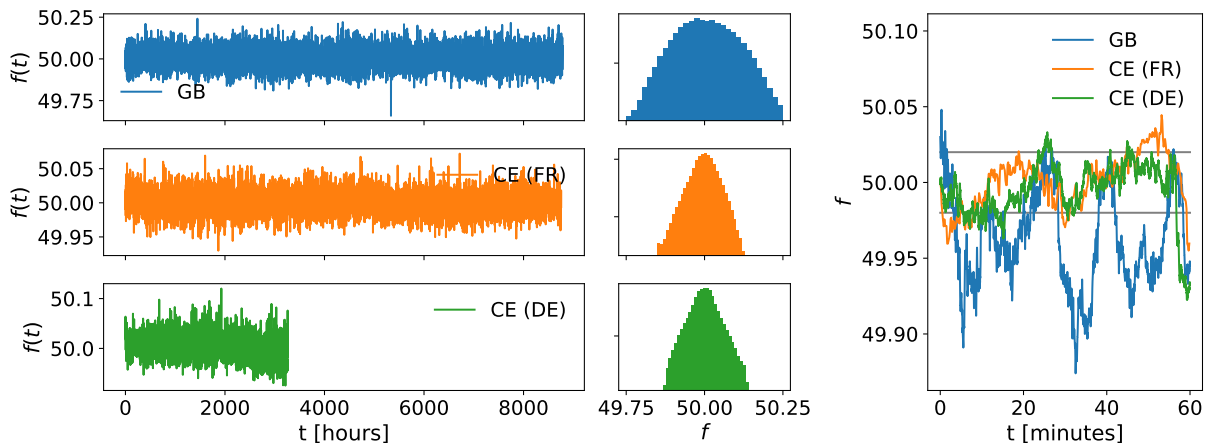
Figure 5.13

Three time series of power grid frequency data: LEFT: plot of one value per hour over the full measurement time for GB (TOP), France (MIDDLE) and Germany (BOTTOM). CENTER: Histograms of all three available time series, using the complete datasets. RIGHT: The measured frequency fluctuations versus time for one hour for both the GB and CE grid. The Horizontal lines indicate the 20mHz interval around 50Hz.

that are also present in the real system. This text is based on the work presented in [I], which follows an entirely data-driven approach. The fluctuation function of the measured frequencies is decomposed in the same way as it was done for temperature data in the previous sections. The inferred characteristic timescales correspond to the processes that generate or control the frequency fluctuations. The work for the article [I] was done in collaboration with Mehrnaz Anvari. All figures are taken from the manuscript.

In the first subsection, the data is shown and the features that are already known are introduced. In the following, the dynamics is analyzed in several steps. First, the typical daily dynamics is presented, then the fluctuations on scales from seconds to days are discussed. Finally, the long time behavior is analyzed.

5.3.1 Frequency data

Recorded frequency data from the power grids in Great Britain (GB) and Continental Europe (CE) is analyzed. The data explanation and interpretation in [I] were done by Mehrnaz Anvari. The time resolution of the data belonging to GB is 1 second, and the data was measured in 2016. For the CE grid, two datasets are considered that were recorded in France in 2015 and Germany 2017. The time resolution of the French dataset is only 10 seconds, while the German dataset has a resolution of 0.2 seconds. However, in contrast to the French and British dataset, the German dataset does not cover a full year. Therefore it can only be used for the short time behavior, while the French data gives insights into the fluctuations for longer times. The available data is presented in figure 5.13.
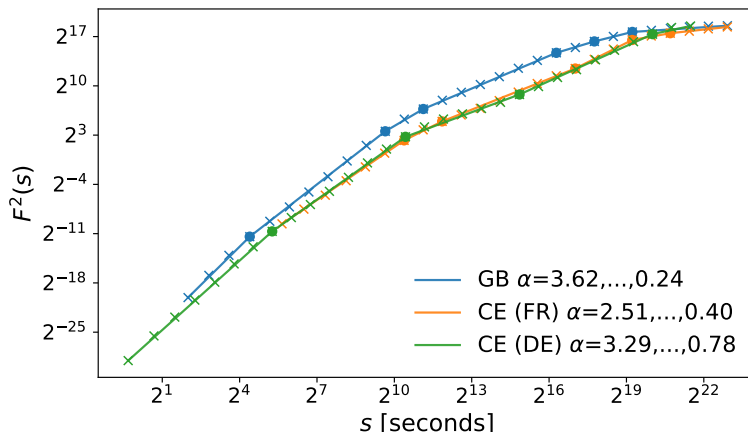
Figure 5.14

DFA of the frequency for the GB grid and the CE grid. The fluctuation functions do not show asymptotic scaling. So in different time–scales, $F^2(s)$ is fitted by $s^{2\alpha}$ with different α values. $F^2(s)$ has a similar slope for both the German ($2\alpha = 3.29, 2.60, 1.36, 1.66, 0.78$) and French ($2\alpha = 2.51, 1.83, 1.46, 1.83, 0.65, 0.40$) dataset as they are parts of the CE power grid [9]. The value of $F^2(s)$ is larger in the GB grid than in the CE grid, as the frequency variations in the GB grid are larger. For GB the slope is described by $2\alpha = 3.62, 2.85, 2.15, 1.55, 1.10, 0.91, 0.24$.

The two power grids in GB and CE differ in their policy. While the operational boundaries in GB are 50Hz $\pm$ 200mHz, in CE they are tighter, namely 50Hz $\pm$ 20mHz [38]. This means that already for smaller deviations from 50Hz, in CE, control systems are activated in order to return to the nominal frequency. The full DFA fluctuation functions for both grids are shown in figure 5.14.

These control systems are important drivers in the investigated data. There are three of them. The primary control works on short timescales of seconds in order to stop further deviation from 50Hz. The secondary control is activated in order to restore the nominal frequency after a few minutes. The tertiary control starts even later after several minutes.

One more important feature of power grid fluctuations is known. This is the 15 minutes period of the energy trading market. Trading is not continuous but happens periodically [116]. The period was already found in power grid frequency time series [111]. Obviously, consumer behavior also exhibits dependence on the day and hour and therefore leads to different characteristics of fluctuations.

The features, mentioned in the last paragraph, are more or less regular. As a first step in the next section, the resulting cycles will be analyzed.

5.3.2 The daily cycle

The energy trading market causes regular patterns in the dynamics of power grids. The European parliament research service describes the characteristic times of energy trading in [116]. They include long time contracts up to 20 years and weekly changes.
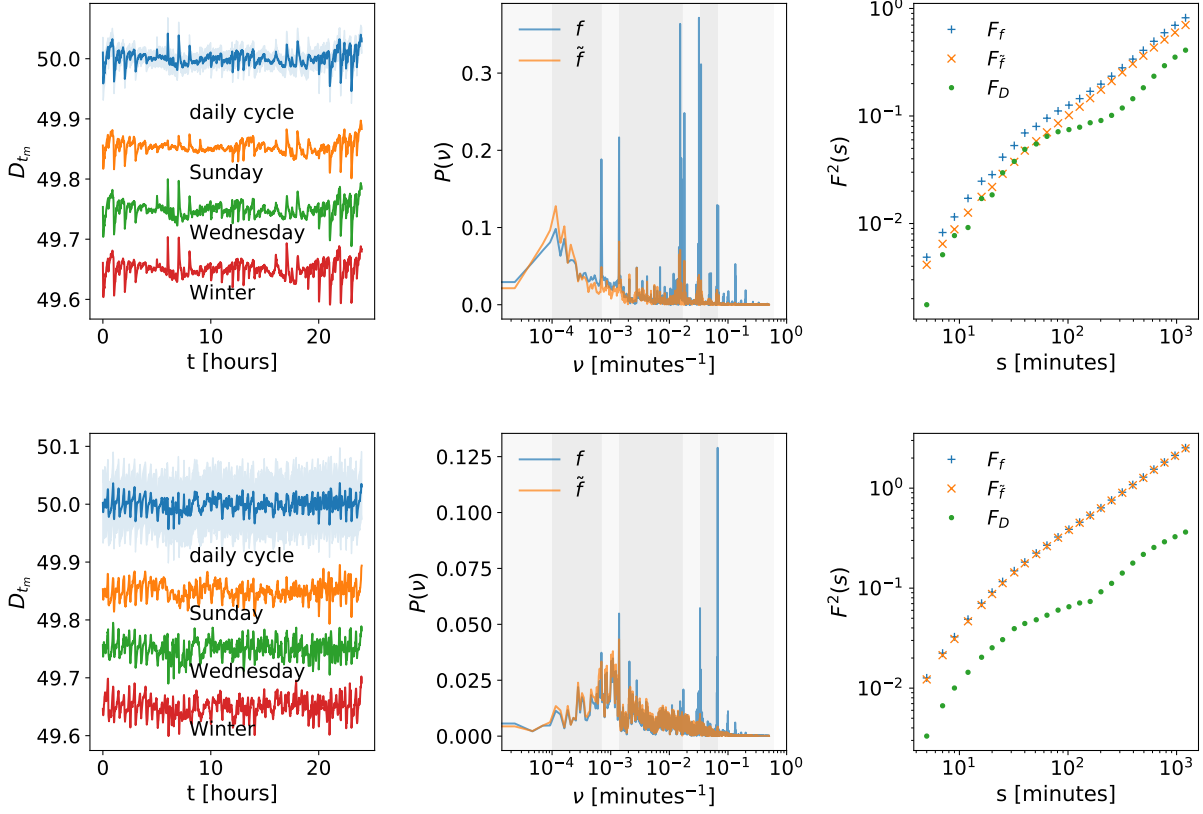
Figure 5.15

LEFT: Average daily cycles of grid frequencies for the CE and the GB grid for all days (blue), Sundays (orange) and Wednesdays (green), representative for weekdays; The average cycle in winter in shown in red. CENTER: Power spectrum before and after removing the daily cycle. RIGHT: DFA fluctuation function before and after removing the daily cycle.

However, the market normally operates on smaller daily and intra-day timescales. The most pronounced peaks in the power spectrum, in the central panels of figure 5.15, are due to those. They are at 15, 30, and 60 minutes, as well as half a day, and one day, highlighted by the stripes in the background of the figure. Every hour, the power plants are dispatched due to the trading market [129]. For the analysis, data from GB and FR with one minute resolution $\lfloor f(t) \rfloor_{1m}$ is used. The power spectra are calculated as the Fourier transforms of the autocorrelation functions $\mathcal{F}[\overline{f(t + \Delta)f(t)}/\overline{f^2(t)}]$.

Since the period 24 hours is a multiple of all observed periods it is sufficient to consider the daily periodicity. It can be detrended in a similar way to the seasonal cycle in climate time series (see section 2.2.2). The anomalies $\tilde{f}$ of the power grid frequency are obtained by calculating the mean frequency for each minute of the day $D(t_m)$

$$\tilde{f}(t_m) = f(t_m) - D(t_m). \tag{5.14}$$

The obtained daily cycles are shown in the left panel of figure 5.15. The daily schedule

is very pronounced for both datasets. For CE, it is more clear due to less noise. For GB, there are more jumps in the daily cycle, however, they are of the same magnitude as the error bars.

Weekdays and Sundays differ, however, the qualitative structure does not change. The weekly cycle, which is, undoubtedly, there is sufficiently small to be ignored for this analysis, at least in GB. For CE it will be discussed later. The seasonal cycle can not be determined due to the lack of interannual data.

After the daily cycle was subtracted, the peaks in the power spectrum in figure 5.15 are significantly suppressed for the time series $\tilde{f}$. The right panel of figure 5.15 shows the effect of subtracting the daily cycle on the DFA fluctuation function $F_{\tilde{f}}(s)$. The effect is very different for CE from GB. While for the CE grid the impact of the daily cycle is strong and $F_{\tilde{f}} \neq F_f$, for GB the fluctuation function hardly changes after removing the cycle.

This is despite the daily cycle having a similar amplitude for GB as it has for CE. The peaks in the power spectrum are clear for both datasets. The reason is the stronger background noise in the GB grid that hides the daily cycle. The operational frequency range for the GB power grid is larger than for CE, i.e. it is 200mHz instead of 20mHz. Here it becomes obvious how much DFA smoothens the data. While the periods are clearly visible in the power spectrum of the GB power grid frequency, they are negligible in DFA.

5.3.3 Analysis of intra-day fluctuations

In this section, the power grid frequency anomalies $\tilde{f}$ of GB and CE are analyzed with the method presented in section 5.1.3. The feature that is expected to be most prominent here is the impact of the control systems.

Control systems are modeled with AR(2) processes. The interpretation of this process now differs a bit from the interpretation in sections 5.1.4 and 5.2.2, where noisy oscillations were described.

Here the noise and the responding system are clearly distinguishable. The noise describes the perturbations of the grid that originate from consumer behavior and variability in renewable energy production as well as the irregularities of the trading market. In an uncontrolled system, such perturbations might lead to diffusive behavior as shown in figure 5.2 by the dashed line. AR(2) acts as a filter that keeps the frequency stationary. While in an uncontrolled system the variance would grow in time, in an AR(2) process $X(t)$ it reaches a stable value σ_X^2. As mentioned in section 5.1.1 the filtering properties might be described as the ratio of the variance of the initial perturbations σ_ϵ^2 and the final variance of the process σ_X^2. The value $\sigma_X^2/\sigma_\epsilon^2$ will always be greater than one since no control system manages to equalize the perturbations (σ_ϵ^2) immediately. Instead, there is always some finite correlation time r.

It is assumed that there are three control systems active in the time range of seconds up to one day. This corresponds to the information given in [38]. The long time behavior is not discussed here but in the following section. The ansatz is, therefore, a superposition
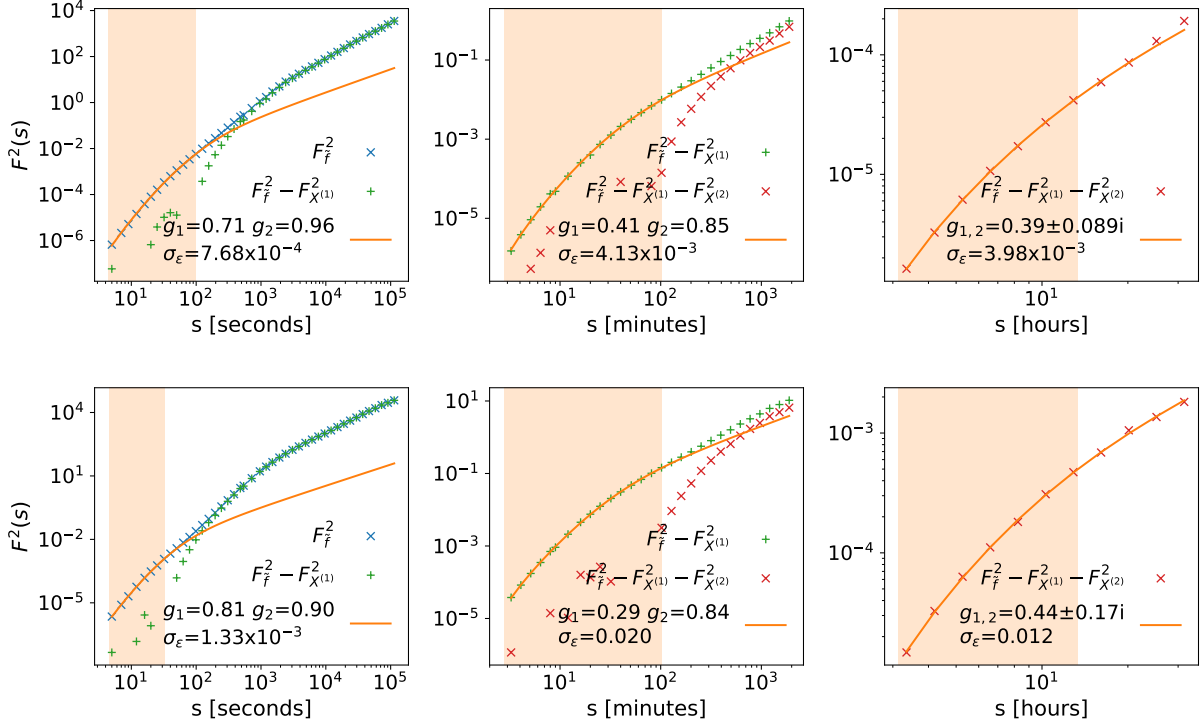
Figure 5.16

Decomposition of fluctuation functions of $\tilde{f}$ for CE (TOP) and GB (BOTTOM). LEFT: Short time fit of an AR(2) model to the fluctuation function (blue 'x'). The fitted curve (orange) is then subtracted. The result $F_{\tilde{f}-X^{(1)}}$ are the green '+'. CENTER: $F_{\tilde{f}-X^{(1)}}$ is rescaled to a one minute resolution and fitted by AR(2) (orange curve). Subtracting the result yields $F_{\tilde{f}-X^{(1)}-X^{(2)}}$ (red 'x'). RIGHT: These dots are rescaled to an hourly timescale and fitted again for longer times by an AR(2) model (orange).

of three processes on different timescales, i.e. seconds, minutes, and hours

$$\tilde{f}_t \approx X^{(1)}(t_s) + X^{(2)}(t_m) + X^{(3)}(t_h). \tag{5.15}$$

Using the superposition principle (4.18) the fluctuations in both the GB and the CE grid can be decomposed. For CE we use data from Germany with one second resolution for short times and data from France with one minute resolution for longer times. In order to match both fluctuation functions in figure 5.16 the values of the fluctuation function for Germany are anomalies $\tilde{f}(t_s)$ just like the data from France.

The fitting procedure is the same as in sections 5.1.4 and 5.2.2. It is shown in figure 5.16. The theoretical fluctuation function F_X of AR(2) are fitted to $F_{\tilde{f}}$ for interval of s highlighted in the background of figure. The three timescales are seconds (left panel), minutes (central panel) and hours (right panel).

The superposition of the three processes explains the fluctuations of the power grid frequency up to the timescale of one day. For each of the three fits, the three parameters

		CE	GB
$r^{(1)}$	[s]	27.0	14.4
$\sigma^{(1)}/\sigma_\epsilon^{(1)}$		11.4	10.1
$\sigma_\epsilon^{(1)}$	$[0.001 s^{-1}]$	0.768	1.33
$r^{(2)}$	[m]	7.29	6.25
$\sigma^{(2)}/\sigma_\epsilon^{(2)}$		3.03	2.46
$\sigma_\epsilon^{(2)}$	$[0.001 m^{-1}]$	4.13	20.4
$r^{(3)}$	[h]	1.08	1.34
$\sigma^{(3)}/\sigma_\epsilon^{(3)}$		1.36	1.48
$\sigma_\epsilon^{(3)}$	$[0.001 h^{-1}]$	3.98	12.0

Table 5.1
Results of the fits of the relaxation times r, of σ_ϵ, and of $\sigma_X^2/\sigma_\epsilon^2$ for all three DFA-fits for both datasets. While the relaxation times are similar for CE and GB, the noise variances σ_ϵ are higher in GB and differ strongly.

to be determined are a, b and σ. The results of all fits are displayed in terms of the observables r, σ_ϵ and σ_X/σ_ϵ in table 5.1. The filtering σ_X/σ_ϵ of the control systems seems to be very similar for GB and CE. The correlation times are also close, at least $r^{(2)}$ and $r^{(3)}$. The main difference seems to be the input noise σ_ϵ on all three timescales. This quantity reflects the policy of when the control system is activated. It is no surprise due to the tighter operational boundaries that the noise variances are lower for the CE power grid frequency data. This is because small fluctuations are already filtered before they might grow larger.

It turns out that, for both the CE and GB grid, the third control system $X^{(3)}$ on the hourly timescale is best described by an AR(2) model with complex roots. So the system overshoots here and, therefore, starts oscillating. The period length of this damped oscillation is approximated by 17 hours for the GB grid and 22 hours for the CE grid.

The three processes $X^{(1)}(t_s)$, $X^{(2)}(t_m)$ and $X^{(3)}(t_h)$ and the daily cycle $D(t_m)$ together represent the characteristic timescales in the power grids up to a day. For short times, the dynamics can be approximated by the short time fit $X^{(1)}(t_s)$ only (at least for GB where the daily cycle is not as important). Short trajectories of the model $X^{(1)}(t_s)$ and the GB data in one second resolution are compared in figure 5.17 in the left panel. On the timescale of minutes, the dynamics of the power grid frequencies $\lfloor f(t_s) \rfloor_m$ is approximated by the superposition

$$\lfloor f(t_s) \rfloor_m \approx \lfloor X^{(1)}(t_s) \rfloor_m + D(t_m) + X^{(2)}(t_m). \tag{5.16}$$

For extending the timescale further and going to one hour time steps, the dynamics of the frequency $\lfloor f(t_s) \rfloor_h$ is approximated by

$$\lfloor f(t_s) \rfloor_m \approx \lfloor X^{(1)}(t_s) \rfloor_h + \lfloor X^{(2)}(t_m) + D(t_m) \rfloor_h + X^{(3)}(t_h) =: Y(t_h). \tag{5.17}$$
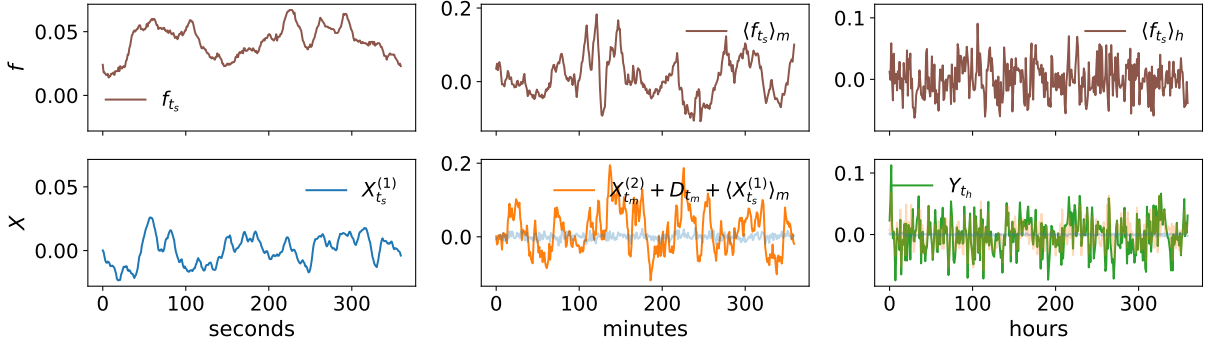
Figure 5.17

Model trajectories (BOTTOM) vs measured frequency data (TOP) from GB for different timescales. LEFT: trajectories f_t with one second resolution compared to a realization of X_t. CENTER: One minute average $\langle f_t \rangle_m$ compared to $\langle X^{(1)}_{t^{(1)}} \rangle_m + X^{(2)}_{t_m} + D_{t_m}$. RIGHT: One hour average $\langle f_t \rangle_h$ compared to $\langle X^{(1)}_{t_s} \rangle_h + \langle X^{(2)}_{t_m} + D_{t_m} \rangle_h + X^{(3)}_{t_h}$.

This full model Y, shown in the right panel of figure 5.17, is the final result of this section. The figure shows that its output looks similar to the measured data. Thus, with the decomposition, a model was obtained that accurately describes the power of the fluctuations in the data from seconds up to a day. Longer timescales are not yet captured by this model and require further analysis in the following section.

5.3.4 Long time regularity

The large s behavior of the power grid frequency fluctuation function (see figure 5.14) is different from all cases discussed so far. The exponent α in $F^2 \propto s^{2\alpha}$ is, for all three time series, smaller than $1/2$. This result implies anti-persistence. It can be modeled, for example, with an ARFIMA(0, d, 0) process with $d < 0$. This is not the same as long range correlations because the cumulative sum of the correlation function for all lag times does not diverge. The power law decay of the correlation function (that also switches signs) is sufficiently steep to ensure convergence [21]. Nevertheless, it is associated with a Hurst exponent. Practically, it means, that the long time average $1/T \sum_{t=0}^{T} f(t)$ stays closer to 50Hz than one would expect from a random process.

How can this feature be explained for the power grid fluctuations? Most likely, the long time regularity is enforced by some policy of the supplies. It is not described by the model Y that describes fluctuations up to one day.

A closer look at the comparison of the one-day-model Y and the measured fluctuation functions in figure 5.18 reveals that there is a clear difference between the two grids. In the GB grid the real fluctuations are clearly smaller than the predicted fluctuations of the model $F_{\tilde{f}} < F_Y$ for large s. At the same time, for CE, it is the other way round and $F_{\tilde{f}} > F_Y$ for large s. So both cases have to be treated differently.

For GB the observed fluctuations can be explained by a simple adjustment of the
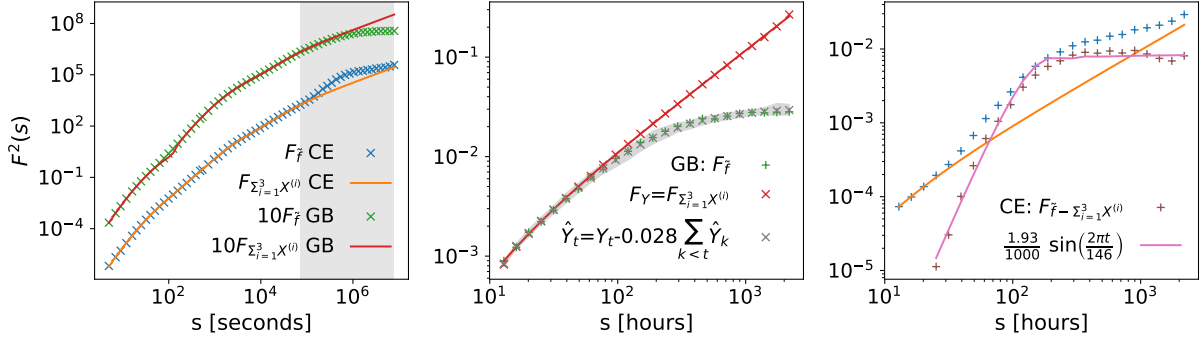
Figure 5.18

Long time behavior of the fluctuation functions. LEFT: Full fluctuation functions of CE and GB and the respective fits of Y. CENTER: $F^2_{f(t_h)} - F^2_{Y(t_h)}$ (brown crosses) fitted with the fluctuation function of an AR(2) process (pink line). The result is a sinusoidal oscillation with period 146 hours. RIGHT: The fluctuation functions of $Y(t_h)$ (red) and $f(t_h)$ (green). The gray dots represent the fit of $\hat{Y}(t_h)$ to $f(t_h)$, i.e. the green points. The shadow represents the deviations in an ensemble of 100 fits.

third control system. In order to keep the long time average stable around 50Hz, the regularization can be done by including a term that sums over all past deviations from 50Hz,

$$\hat{Y}(t_h) = Y(t_h) - Q \sum_{\Delta=0}^{t} Y(\Delta_h), \tag{5.18}$$

where $Y(t_h)$ is the result of equation (5.17) and $\hat{Y}(t_h)$ is the final model. This extension of the model leads to one additional free parameter Q, which is fitted in figure 5.18 in the central panel for an ensemble of realizations of Y. The result is $Q = 0.028 \pm 0.003$. The fluctuation function of the modified model nicely matches the measured fluctuation function for large s.

The fluctuation function of the power grid frequency of the CE grid can not be described by Y with simply adding a filter. In fact, the real system fluctuates stronger than the model. So a process on a long timescale has to be added. In the right panel of figure 5.18, the fluctuation function of $\tilde{f} - X^{(1)} - X^{(2)} - X^{(3)}$ on the timescale of several hours up to months is fitted by the AR(2) model. The fit yields a sinusoidal behavior with a period of 146 hours (≈ 6.2 days). The deviations between F_{data} and F_{model} suggest that this model $\hat{Y}$, despite being the best fit, does not accurately describe the dynamics. In the power spectrum, in figure 5.15 (top center), there is also no clear peak visible that would indicate a sinusoidal oscillation. However, there is a broad local maximum at the timescale in question so, apparently, something is happening there.

Since the period is close to a week, the feature might originate from weekly policy changes or trading. It follows, however, not a pure oscillatory dynamics but something that can not be accurately modeled with the introduced methods. Nevertheless, adding

the fitted process to the model gives a nice and simple data model for the CE power grid fluctuations. It explains why the fluctuations in the long time limit in figure 5.14 are almost identical to the GB grid fluctuations, even though the short time fluctuations are more suppressed in CE.

In conclusion, for both systems the GB grid and the CE grid a suitable data model was created. The model for GB includes a daily cycle plus 10 free parameters and explains the observations with three AR(2) filters and an enforced long time regularity. The AR(2) filters correspond to the three control systems in the power grid. The model for CE includes a daily cycle plus 11 free parameters in three AR(2) filters and one long time oscillation that is known to be a rather inaccurate description. The parameters can be compared, which makes it possible to trace back the differences and show similarities between the two systems.

6 Conclusions

Accurate modeling of many physical processes is very complicated due to the high complexity of the systems. In these cases, there is a demand for data models that reproduce the statistical properties of the system without requiring deep physical insight into the dynamics.

This thesis covers different systems and scenarios. All discussed topics have in common that they deal with the categorization and decomposition of fluctuations in one-dimensional time series. The methods enable the user to generate an accurate data model and describe the properties of the system such as the predictability. This is done for many different types of dynamics in a very general setting. Consequently, the methodology can be applied to very different types of complex systems like power grid fluctuations or climate variability. Both of the aforementioned examples were discussed here in more detail.

Understanding of detrended fluctuation analysis has improved in recent years. This work draws several new conclusions from the analytical description developed by Marc Höll [56]. The results should further be relevant for the scaling analysis based on the wavelet transform and detrended moving averages. By digging deep into DFA, other properties apart from long range correlations can be inferred. It is shown that from the DFA fluctuation function, characteristic timescales, oscillations, and relaxations can be found, even for complex processes with several timescales.

The second topic addressed in this thesis is the origin of long range correlations in data. This is discussed in different sections throughout the text.

In a more general setting, a framework of anomalous statistical effects is discussed. There are three anomalous statistical effects, which violate the premises of the central limit theorem and lead to the so-called Hurst effect and anomalous diffusion. According to a recently proposed framework, there are three possible causes. The Joseph, Noah, and Moses effects can be described by exponents that are directly related to the Hurst exponent by a scaling relation. In this work, this framework is further investigated and applied to deterministic anomalous diffusion.

6.1 Decomposing Hurst exponents

Anomalous statistical properties are observed in several systems. They are detected with several methods that are not in all cases equivalent. It is important to distinguish different types of anomalous properties in order to find an adequate model for the data.

A recently proposed framework introduces three root causes for anomalous diffusion, quantified by the Hurst exponent H. These are the Joseph effect, the Noah effect, and

the Moses effect. Here the Moses effect is a new ingredient of [27], while the Joseph and Noah effects were coined by Mandelbrot.

The three exponents describing the effects were previously defined by the result of R/S statistics, the scaling of the absolute value of the first moment and the scaling of the second moment. In this work, the framework was linked to a rigorous derivation of the ensemble averaged time averaged mean squared displacement. By doing so, for a large class of systems, the direct relation of the Hurst exponent to the exponents that define the effects was proven.

Each of the three effects corresponds to a paradigmatic model that is simple and well known in the literature. Theses models are fractional Brownian motion, scaled Brownian motion, and motion with Lévy increments. The result for each of these models gives one pure cause of anomalous diffusion.

An effect that was not considered before in this context is anomalous diffusion due to regular periods with diverging mean duration. This is known to lead to aging and infinite densities. The Pomeau-Manneville map is a deterministic system that exhibits this behavior. The employed version of the Pomeau-Manneville map is symmetric and its iterates can, therefore, be regarded as the increments of a process that exhibits anomalous diffusion. Here, depending on the parameter choice, the decomposition yields complex combinations of all three, the Joseph, Noah and Moses effect. The system can be mapped to a stochastic renewal process and thereby the scaling of the mean squared displacement can be found. The Noah effect and the Moses effect can be calculated from the infinite densities of the map that were previously derived.

Since the system is aging, a Moses effect is expected. However, the other results are surprising at first sight. Depending on the parameter the system also exhibits a Joseph effect and a Noah effect. The Joseph effect can be explained by the diverging mean waiting time in the regular region. The Noah effect is an artifact of the aging. When the variable is rescaled to compensate for the Moses effect, the distribution becomes fat tailed, even though the physical density is bounded.

The decomposition was discussed for a number of additional systems. It was concluded in section 3.3.3 that the Joseph effect is never present for jump processes. In section 3.3.1 the characteristic combination of Noah and Moses effect in the presence of an infinite invariant density was derived. Moreover, the case of extremely fat tails was discussed. They lead to aging and, therefore, to a Moses effect in addition to a maximal Noah effect. So, on the one hand, it is possible that aging gives rise to a Noah effect. On the other hand, very fat tails of the increment distribution can lead to a Moses effect. Both do not contradict the interpretation of the decomposition. The Moses effect indicates aging, which is indeed present for very fat tailed distributions. The Noah effect, in fact, describes the tails for the stationary (rescaled) distribution. These results can help to interpret observations of complex combinations of the Joseph, Noah, and Moses effects in measured data or for models that are not analytically solved. Understanding the method is essential for applications in such cases.

It would be interesting to similarly decompose the anomalous diffusive behavior found in other systems, especially experimental systems. Here one problem is that in many cases only one realization of the system exists. Due to weak ergodicity breaking, some-

times a time average can not be used for calculating mean values as they are defined in the framework. However, it was shown in section 3.3.3, that at least the Joseph exponent seems to be stable despite the values of the time average being ill-defined. Also for some systems, there may be detectable intermittent triggering events that are renewals in a stochastic sense. They can, therefore, be treated as the beginning of a new process [16, 27]. In this way, an ensemble can be created.

Possible applications for the theory are in all systems that are suspected to yield anomalous diffusion or anomalous statistical properties. The framework provides a fast and intuitive first understanding of the nature of the fluctuations. It can be used for interpretations and model selection.

6.2 Dynamical causes for long range correlations

By looking at the three effects described above, one can ask when they appear and which the possible dynamical origins are. With the examples presented throughout this thesis, this can be answered for the Joseph effect especially well.

Long range correlations were detected in many different systems. The paradigmatic model for the Joseph effect is fractional Brownian motion. This is a process driven by fractional Gaussian noise. In its discrete version, the autoregressive fractionally integrated moving average process can be used, which was introduced in section 2.2.4. The dynamics of this model depends on all its past increments. It can be argued whether or not this is a situation expected to be physical. For the atmosphere, the presence of such a memory kernel would imply an infinite number of layers, e.g. in the ocean, that memorize the past states of the system.

A second dynamical origin of real long range correlations was presented in section 3.2. The Pomeau-Manneville map is a well-known model of anomalous diffusion. However, since its correlation function is not known, it is not obvious how to link it to long range correlations. This is enabled by the framework of decomposing Hurst exponents. Indeed, for the parameter $z > 2.5$, the Joseph exponent yields long range correlations. The same route to long range correlations is realized in the corresponding stochastic models, i.e. velocity renewal processes. Here, long range correlations can be found even for finite mean waiting times if the jumps sizes grow with long waiting times.

Besides the dynamical models that show real long memory, the detected correlations in data can also be explained by short range correlated models. This is always true because long range correlations as an asymptotic property can not be proven with a finite time series. Detrended fluctuation analysis as a popular tool can only examine timescales up to one fourth of the length of the time series. Especially in climate science, where good datasets cover only around 100 years of data, this is a problem.

As an example, the fluctuations in global mean surface temperature measurements were analyzed in section 4.2. In DFA they are long range correlated. The data was previously modeled by a zero-dimensional energy balance model containing the El Niño southern oscillation and volcanic eruptions as the main causes of variability. This model exhibits AR(1)-type short range correlations in the considered time range. When the

model is fitted to yearly data it matches fairly well. It was shown that the fluctuations on the longest available timescale for both the data and the model yield similar DFA fluctuation functions. Therefore, the model is able to explain the observed long range correlations. In order to see the short range correlated asymptotics in DFA, 1000 data points are needed for the model. This corresponds to 1000 years of temperature recordings. Today, only around 150 years are available.

Obviously, not all datasets where long range correlations were detected can simply be described by exponentially decaying correlations. However, there is a second scenario of how apparent long range correlations can emerge without a single dynamical cause. This happens when several processes with different characteristic time scales are put together and create fluctuations on multiple time scales.

As an example of how this mechanism might work, temperature data from European stations was analyzed in section 5.2.2. The fluctuations were previously described by a long range correlated ARFIMA model. It turns out that, for many stations, they can as well be modeled by a superposition of three simple short range correlated processes. These are autoregressive models and therefore low pass filters. This means that they 'hide' the fluctuations of their drivers on short timescales. Superimposing autoregressive processes with different periods leads to a successive increase in the background noise for long times. The three processes are: exponential decay of correlations for short times, the seasonal cycle, and a 7–8 year noisy oscillation. All of these constituents were previously detected with other methods.

So this thesis identifies four causes of measuring long range correlations. The phenomenon is present in

- dynamical long range correlations due to a memory kernel like in the paradigmatic fractional Gaussian noise model,

- dynamical long range correlations due to intermittency, where the regular periods or the displacement during the regular period have diverging mean values,

- short range correlated processes, if the correlation time is long compared to the length of the time series (this is already the case for correlation times around two it less than 1000 data points are available), and

- emergent scaling of fluctuations due to a superposition of short range correlated processes in the time range under consideration.

The results from section 5.2 imply for temperature variability that the fourth explanation is most plausible since decomposing the fluctuation function yields a model that accounts for the full power of the fluctuations and reproduces characteristics of the data that were previously found with different methods. For future work, a similar critical discussion of the dynamical origins would be interesting for other datasets in which long range correlations were detected.

The existence of 'real' dynamical long range correlations can never be proven or disproven with a finite dataset. However, for modeling, it is also not always relevant. As

long as the time for which the model is supposed to make predictions is shorter than the analyzed time range, the asymptotic behavior of the model is not important. Implications of long range correlations, like effective shrinking of the sample size, still hold even if the observed scaling is just due to a superposition of some short range correlated processes.

6.3 Inferring characteristic timescales

In order to be able to discuss reasons for emerging long range correlations that do not have a true dynamical origin, a deep understanding of the detection method is necessary. In this thesis, the results of detrended fluctuation analysis were discussed. Here analytical understanding has recently improved. The relation between the autocorrelation function and the DFA fluctuation function is known. Consequently, the expected fluctuation function of different model processes can be calculated and compared to empirical results. This was done for global mean surface temperatures in section 4.2. In this way, the theory of DFA can be used for validation of results.

A more interesting application is using it for inferring properties of the data and generating data models. It was shown in this text that this idea works for short range correlated processes, as well as for long range correlated processes, where it is traditionally used. This approach is new and was published in [F]. In section 5.1.3 an extension of the method was presented based on the superposition principle of DFA. It enables the user to investigate time series with more than one characteristic timescale. These characteristic timescales have to be well separated. They are described by models, which are low pass filters. In this thesis, the features were either described by an exponential relaxation (an AR(1) model) or an AR(2) model that depending on the parameters exhibits relaxations and/or oscillations. The noise distribution of the inputs does not have to be Gaussian since the method only reproduced the autocorrelations. In this sense, AR(1) and AR(2) can be generalized when results are interpreted.

Several measured time series were investigated as examples. These were mostly taken from atmospheric science. For instance, the method was applied for estimating the correlation time of global mean surface temperatures, which yields approximately 10 months. It was also shown that the method leads to inconsistencies when applied to monthly data due to interannual variability.

For European station data, sea level pressure anomalies are well approximated by AR(1) models, i.e. exponential decay of correlations. These time series yield relaxation times of 3–4 days. In contrast to this result, a recorded series from the equatorial pacific area contains two pronounced oscillatory modes around 40 days and 40 months. These are manifestations of the well known Madden-Julian oscillation and the El Niño southern oscillation respectively. The latter was further investigated by looking at two measures that are designed to describe the phenomenon. Monthly values can not be described by a plain noisy oscillation (an AR(2) model) due to short time variability. However, when averaging over eleven months the averaged time series is well described by an AR(2) model with a noisy oscillation. The period of the oscillation is indeed around 40 months.

So this same period was found in several datasets with two different approaches, which raises confidence that it is, in fact, a physical quantity. In the literature, the exact nature of the El Niño southern oscillation is still debated. Its interpretation as an oscillatory mode is not generally accepted. Instead, it might also be a two-level system driven by a stochastic input [128].

The method of fitting fluctuation functions is not only able to detect the most prominent oscillatory mode in climate time series, but also more hidden ones. The 7–8 year cycle in European temperature variability was previously detected with spectral methods. In section 5.2 it was shown that it can also be detected by fitting fluctuation functions when decomposing European temperature variability. This yields a superposition of short time variability, the seasonal cycle, and some additional interannual variability. For many stations, this interannual variability is best described by a noisy oscillation of the period of 7–8 years. The method can distinguish datasets where the cycle is significant to those where it is not present or negligible. The findings explain the emerging long range correlations, that are observed in temperature time series.

As a last example, the fluctuation functions of power grid frequencies for two European grids are decomposed using the same method. Even though the nature of the fluctuations is different here and mostly reflects human decisions and supplier-policies, the approach still leads to a meaningful result. The three control systems in the power grids could be identified and compared. It was also shown that the long time variability is less pronounced than one would expect from a noisy system indicating some additional regulation.

The method of fitting DFA fluctuation functions was therefore shown to be suitable for tasks that are traditionally done with several other methods. One is direct fitting of autoregressive models. Here only the very short time behavior is taken into account. Relaxations can also be found by looking at the autocorrelation function and oscillations (in addition to relaxations) can be found by looking at the power spectrum. Besides these standard approaches, there are more sophisticated methods like singular spectrum analysis [99], or even more advanced versions of it [96]. Out of many others, the different implementations of wavelet methods [1, 12] should be noted here, which are closely related to DFA [57].

After applying the method of fitting DFA fluctuation functions to a large number of datasets, the method has turned out to be valuable for several reasons. The advantages are

- immediate validation of the result; the error is given by the deviation of the fluctuation function of the model to the fluctuation function of the data,

- that characteristic timescales are ordered and can, therefore, be investigated sequentially,

- stability of the fluctuation function up to one fourth of the total length of the dataset; therefore long times can be taken into account compared to other methods,

- the possibility of simultaneously finding relaxations and oscillations in the data even if they are hidden under white noise,

- results that yield a complete linear data model that accounts for the complete power of the oscillations, and

- the possibility of neglecting polynomial trends, i.e. slow dynamics, by choosing suitable detrending orders in DFA.

It can be concluded, that the method of fitting fluctuation functions is a useful new tool in the box of time series analysis. DFA has advantages over other methods, especially for short datasets or datasets with trends, where direct measurement of the correlation function is hopeless. It also has advantages over spectral methods, namely that it accounts for the full power of the fluctuations and provides a complete model.

Of course, the main advantage of DFA, namely its smoothness, might also be a disadvantage when details of the dynamics are of interest and not only the dominant timescale. Those are often not visible due to the superposition principle of DFA. The method also does not provide any information about the probability density of the process. It neither captures non-linearities nor interactions between dynamics on different timescales. Also, for accurate measurements of oscillation periods, spectral methods seem to be more suitable than fitting the fluctuation function. At least, the obtained uncertainty of 1.8 degrees for the 7–8 year period is well above uncertainties in previous studies. Still, understanding fluctuation functions has turned out to be useful for inferring the dominant features of systems and building simple data models. Due to the generality of the problem of detecting characteristic timescales, these findings can be relevant for many different fields of research other than atmospheric science and power grid engineering.

Bibliography

[1] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory*, 44(1):2–15, Jan 1998.

[2] E. Aghion, D. A. Kessler, and E. Barkai. From non-normalizable boltzmann-gibbs statistics to infinite-ergodic theory. *Physical review letters*, 122(1):010601, 2019.

[3] T. Akimoto and E. Barkai. Aging generates regular motion in weakly chaotic systems. *Phys. Rev. E*, 87:032915, 2013.

[4] T. Akimoto, E. Barkai, and G. Radons. Infinite invariant density in a semi-markov process with continuous state variables. *arXiv preprint arXiv:1908.10501*, 2019.

[5] T. Akimoto and T. Miyaguchi. Phase diagram in stored-energy-driven levy flight. *J. Stat. Phys.*, 157:515, 2014.

[6] T. Akimoto, S. Shinkai, and Y. Aizawa. Distributional behavior of time averages of non-l1 observables in one-dimensional intermittent maps with infinite invariant measures. *J. Stat. Phys.*, 158:476, 2014.

[7] T. Albers and G. Radons. Exact results for the nonergodicity of d-dimensional generalized levy walks. *Phys. Rev. Lett.*, 120:104501, 2018.

[8] E. Alessio, A. Carbone, G. Castelli, and V. Frappietro. Second-order moving average and scaling of stochastic time series. *The European Physical Journal B - Condensed Matter and Complex Systems*, 27(2):197–200, May 2002.

[9] M. Anvari, L. R. Gorjão, M. Timme, B. Schäfer, D. Witthaut, and H. Kantz. Stochastic analysis of power grid frequency and of a synthetic model. *arXiv preprint arXiv:*, 2019.

[10] S. Auer, F. Hellmann, M. Krause, and J. Kurths. Stability of synchrony against local intermittent fluctuations in tree-like power grids. *Chaos*, 27:127003, 2017.

[11] R. T. Baillie. Long memory processes and fractional integration in econometrics. *Journal of econometrics*, 73(1):5–59, 1996.

[12] S. Baliunas, P. Frick, D. Sokoloff, and W. Soon. Time scales and trends in the central england temperature data (1659–1990): A wavelet analysis. *Geophysical Research Letters*, 24(11):1351–1354, 1997.

[13] J.-M. Bardet and I. Kammoun. Asymptotic properties of the detrended fluctuation analysis of long-range-dependent processes. *IEEE Transactions on information theory*, 54(5):2041–2052, 2008.

[14] F. Bardou, J. Bouchaud, O. Emile, A. Aspect, and C. Cohen-Tannoudji. Subrecoil laser cooling and lévy flights. *Physical review letters*, 72(2):203, 1994.

[15] E. Barkai. Aging in subdiffusion generated by a deterministic dynamical system. *Phys. Rev. Lett.*, 90:104101, Mar 2003.

[16] K. E. Bassler, J. L. McCauley, and G. H. Gunaratne. Nonstationary increments, scaling distributions, and variable diffusion processes in financial markets. *Proceedings of the National Academy of Sciences*, 104(44):17287–17290, 2007.

[17] C. Beck and G. Roepstorff. From dynamical systems to the langevin equation. *Physica A: Statistical Mechanics and its Applications*, 145(1):1 – 14, 1987.

[18] G. Bel and E. Barkai. Ergodicity breaking in a deterministic system. *Europhys. Lett.*, 74:15, 2006.

[19] S. Bo, F. Schmidt, R. Eichhorn, and G. Volpe. Measurement of anomalous diffusion using recurrent neural networks. *arXiv preprint arXiv:1905.02038*, 2019.

[20] M. Bothe, F. Sagues, and I. Sokolov. Mean squared displacement in a generalized lévy walk model. *arXiv preprint arXiv:1903.09505*, 2019.

[21] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time series analysis: forecasting and control.* John Wiley & Sons, 2015.

[22] X. Brokmann, J.-P. Hermier, G. Messin, P. Desbiolles, J.-P. Bouchaud, and M. Dahan. Statistical aging and nonergodicity in the fluorescence of single nanocrystals. *Physical review letters*, 90(12):120601, 2003.

[23] R. Bryce and K. Sprague. Revisiting detrended fluctuation analysis. *Scientific reports*, 2:315, 2012.

[24] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J.-H. Peter, and K. Voigt. Correlated and uncorrelated regions in heart-rate fluctuations during sleep. *Phys. Rev. Lett.*, 85:3736–3739, Oct 2000.

[25] A. V. Chechkin, R. Metzler, J. Klafter, V. Y. Gonchar, et al. Introduction to the theory of lévy flights. *Anomalous Transport*, pages 129–162, 2008.

[26] M. D. Chekroun, D. Kondrashov, and M. Ghil. Predicting stochastic systems by noise sampling, and application to the el niño-southern oscillation. *Proceedings of the National Academy of Sciences*, 108(29):11766–11771, 2011.

[27] L. Chen, K. E. Bassler, J. L. McCauley, and G. H. Gunaratne. Anomalous scaling of stochastic processes and the moses effect. *Phys. Rev. E*, 95:042141, Apr 2017.

[28] Z. Chen, K. Hu, P. Carpena, P. Bernaola-Galvan, H. E. Stanley, and P. C. Ivanov. Effect of nonlinear filters on detrended fluctuation analysis. *Physical Review E*, 71(1):011104, 2005.

[29] N. Crato, R. Linhares, and S. Lopes. Statistical properties of detrended fluctuation analysis. *Journal of Statistical Computation and Simulation*, 80(6):625–641, 2010.

[30] Z. Czechowski and L. Telesca. Detrended fluctuation analysis of the ornstein-uhlenbeck process: Stationarity versus nonstationarity. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(11):113109, 2016.

[31] A. Dechant, E. Lutz, D. Kessler, and E. Barkai. Scaling green-kubo relation and application to three aging systems. *Physical Review X*, 4:011022, 2014.

[32] Q. Deng and Z. Fu. Comparison of methods for extracting annual cycle with changing amplitude in climate series. *Climate Dynamics*, 52(7-8):5059–5070, 2019.

[33] Q. Deng, D. Nian, and Z. Fu. The impact of inter-annual variability of annual cycle on long-term persistence of surface air temperature in long historical records. *Climate Dyn.*, 50:1091, 2018.

[34] N. Dunstone, D. Smith, A. Scaife, L. Hermanson, R. Eade, N. Robinson, M. Andrews, and J. Knight. Skilful predictions of the winter north atlantic oscillation one year ahead. *Nature Geoscience*, 9(11):809, 2016.

[35] E. B. Dynkin. *Selected Translations in Mathematical Statistics and Probability*, 1:171, 1961.

[36] J. Echeverrıa, M. Woolfson, J. Crowe, B. Hayes-Gill, G. Croaker, and H. Vyas. Interpretation of heart rate variability via detrended fluctuation analysis and $\alpha\beta$ filter. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 13(2):467–475, 2003.

[37] A. Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 322(8):549–560, 1905.

[38] ENTSO-E. Statistical factsheet 2014. https://www.entsoe.eu/publications/major-publications/Pages/default.aspx.

[39] K. Fraedrich and R. Blender. Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys. Rev. Lett.*, 90:108501, 2003.

[40] H.-B. Fredriksen and K. Rypdal. Spectral characteristics of instrumental and climate model surface temperatures. *Journal of Climate*, 29(4):1253–1268, 2016.

[41] H.-B. Fredriksen and M. Rypdal. Long-range persistence in global surface temperatures explained by linear multibox energy balance models. *Journal of Climate*, 30(18):7157–7168, 2017.

[42] C. W. Gardiner. *Handbook of stochastic methods*, volume 3. springer Berlin, 1985.

[43] T. Geisel and S. Thomae. Anomalous diffusion in intermittent chaotic systems. *Phys. Rev. Lett.*, 52:1936, 1984.

[44] M. Ghil, M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, et al. Advanced spectral methods for climatic time series. *Reviews of geophysics*, 40(1):3–1, 2002.

[45] M. Ghil, A. Groth, D. Kondrashov, and A. W. Robertson. Extratropical subseasonal to seasonal oscillations and multiple regimes: The dynamical systems view. In A. W. Robertson and F. Vitart, editors, *Sub-Seasonal to Seasonal Prediction*, pages 119 – 142. Elsevier, 2019.

[46] L. R. Gorjão, M. Anvari, H. Kantz, D. Witthaut, B. Schäfer, and M. Timme. Modelling the dynamics of power grid frequency fluctuations from data. *arXiv preprint arXiv:*, 2019.

[47] R. B. Govindan, D. Vyushin, A. Bunde, S. Brenner, S. Havlin, and H.-J. Schellnhuber. Global climate models violate scaling of the observed atmospheric variability. *Phys. Rev. Lett.*, 89:028501, 2002.

[48] C. W. Granger and R. Joyeux. An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1):15–29, 1980.

[49] T. Graves, R. Gramacy, N. Watkins, and C. Franzke. A brief history of long memory: Hurst, mandelbrot and the road to arfima, 1951-1980. *Entropy*, 19(9):437, 2017.

[50] J. Grieser, S. Trömel, and C.-D. Schönwiese. Statistical time series decomposition into significant components and application to european temperature. *Theoretical and applied climatology*, 71(3-4):171–183, 2002.

[51] J. Hansen, M. Sato, P. Kharecha, and K. von Schuckmann. Earth's energy imbalance and implications. *Atmos. Chem. Phys.*, 11:13421, 2011.

[52] K. Hasselmann. Stochastic climate models: Part i. theory. *Tellus*, 28:473–485, 1976.

[53] C. Heneghan and G. McDarby. Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. *Physical review E*, 62(5):6103, 2000.

[54] M. Hoell and H. Kantz. The fluctuation function of the detrended fluctuation analysis – investigation on the ar(1) process. *Eur. Phys. J. B*, 88:126, 2015.

[55] M. Hoell and H. Kantz. The relationship between the detrendend fluctuation analysis and the autocorrelation function of a signal. *Eur. Phys. J. B*, 88:327, 2015.

[56] M. Höll. *Analytical Investigation of Long-range Correlated Time Series and of the Method of Detrended Fluctuation Analysis*. PhD thesis, Technische Universität Dresden, 2018.

[57] M. Höll, K. Kiyono, and H. Kantz. Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average. *Phys. Rev. E*, 99:033305, Mar 2019.

[58] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley. Effect of trends on detrended fluctuation analysis. *Phys. Rev. E*, 64:011114, 2001.

[59] S. Hu and A. V. Fedorov. The extreme el nino of 2015-2016 and the end of global warming hiatus. *Geophys. Res. Lett.*, 44:3816–3824, 2017.

[60] H. E. Hurst. Long-term storage capacity of reservoirs. *Am. Soc. Civil Eng.*, 116:770, 1951.

[61] P. Huybers and W. Curry. Links between annual, milankovitch and continuum temperature variability. *Nature*, 441(7091):329, 2006.

[62] N. Jajcay, J. Hlinka, S. Kravtsov, A. A. Tsonis, and M. Paluš. Time scales of the european surface air temperature variability: The role of the 7–8 year cycle. *Geophysical Research Letters*, 43(2):902–909, 2016.

[63] J.-H. Jeon, A. V. Chechkin, and R. Metzler. Scaled brownian motion: a paradoxical process with a time dependent diffusivity for the description of anomalous diffusion. *Physical Chemistry Chemical Physics*, 16(30):15811–15817, 2014.

[64] L. Jiang, N. Li, and X. Zhao. Scaling behaviors of precipitation over china. *Theoretical and applied climatology*, 128(1-2):63–70, 2017.

[65] J. W. Kantelhardt. Fractal and multifractal time series. *Encyclopedia of Complexity and Systems Science*, pages 3754–3779, 2009.

[66] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde. Detecting long-range correlations with detrended fuctuation analysis. *Physica A*, 295:441, 2001.

[67] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, 2002.

[68] H. Kantz and T. Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge university press, 2004.

[69] J. J. Kennedy, N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker. Reassessing biases and other uncertainties in sea-surface temperature observations measured in situ since 1850 part 2: biases and homogenisation. *J. Geophys. Res.*, 116:D14104, 2011.

[70] R. Klages. From deterministic chaos to anomalous diffusion. *Reviews of nonlinear dynamics and complexity*, 3:169–227, 2010.

[71] R. Kubo. The fluctuation-dissipation theorem. *Reports on progress in physics*, 29(1):255, 1966.

[72] P. Kundur. *Power System Stability And Control*. EPRI power system engineering series. McGraw-Hill, 1994.

[73] S. Lovejoy. A voyage through scales, a missing quadrillion and why the climate is not what you expect. *Climate Dynamics*, 44(11-12):3187–3210, 2015.

[74] S. Lovejoy. Spectra, intermittency, and extremes of weather, macroweather and climate. *Scientific reports*, 8(1):12697, 2018.

[75] J. Ludescher, A. Bunde, C. L. Franzke, and H. J. Schellnhuber. Long-term persistence enhances uncertainty about anthropogenic warming of antarctica. *Climate dynamics*, 46(1-2):263–271, 2016.

[76] R. A. Madden and P. R. Julian. Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *Journal of the atmospheric sciences*, 28(5):702–708, 1971.

[77] B. B. Mandelbrot and R. L. Hudson. *The (mis) behaviour of markets: a fractal view of risk, ruin and reward*. Profile books, 2010.

[78] B. B. Mandelbrot and J. W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[79] B. B. Mandelbrot and J. R. Wallis. Noah, joseph, and operational hydrology. *Water Resources Research*, 4(5):909–918.

[80] D. Manik, M. Rohden, H. Ronellenfitsch, X. Zhang, S. Hallerberg, D. Witthaut, and M. Timme. Network susceptibilities: Theory and applications. *Physical Review E*, 95(1):012319, 2017.

[81] M. E. Mann. On long range dependence in global surface temperature series. *Climate Change*, 107:267–276, 2011.

[82] D. Maraun, H. W. Rust, and J. Timmer. Tempting long-memory – on the interpretation of dfa results. *Nonlin. Process. Geophys.*, 11:495, 2004.

[83] M. Massah. *Convergence of Large Deviations Probabilities for Processes with Memory - Models and Data Study*. PhD thesis, Technische Universität Dresden, 2018.

[84] M. Massah and H. Kantz. Confidence intervals for time averages in the presence of long-range correlations, a case study on earth surface temperature anomalies. *Geophys. Res. Lett.*, 43:9243, 2016.

[85] M. Massah, M. Nicol, and H. Kantz. Large-deviation probabilities for correlated gaussian processes and intermittent dynamical systems. *Physical Review E*, 97(5):052147, 2018.

[86] R. Metzler, J.-H. Jeon, A. G. Cherstvy, and E. Barkai. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Physical Chemistry Chemical Physics*, 16(44):24128–24164, 2014.

[87] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339:1–77, 2000.

[88] P. Meyer. *Anomalous Statistical Properties caused by Dynamical Intermittency*. PhD thesis, Technische Universität Dresden, 2016.

[89] J. M. Mitchell. An overview of climatic variability and its causal mechanisms. *Quaternary Research*, 6(4):481–493, 1976.

[90] E. W. Montroll. Random walks on lattices. iii. calculation of first-passage times with application to exciton trapping on photosynthetic units. *Journal of Mathematical Physics*, 10(4):753–765, 1969.

[91] W. Moon, S. Agarwal, and J. S. Wettlaufer. Intrinsic pink-noise multidecadal global climate dynamics mode. *Physical review letters*, 121(10):108701, 2018.

[92] C. P. Morice, J. J. Kennedy, N. A. Rayner, and P. D. Jones. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the hadcrut4 dataset. *J. Geophys. Res.*, 117:D08101, 2012.

[93] M. Niemann. From anomalous deterministic diffusion to the continuous-time random walk. *Ph.D. diss.*, 2009.

[94] P. O'Connell. A simple stochastic modelling of hurst's law. *Mathematical models in hydrology*, 1, 1974.

[95] M. Paluš. Multiscale atmospheric dynamics: cross-frequency phase-amplitude coupling in the air temperature. *Physical review letters*, 112(7):078702, 2014.

[96] M. Paluš and D. Novotná. Enhanced monte carlo singular system analysis and detection of period 7.8 years oscillatory modes in the monthly nao index and temperature records. *Nonlinear Processes in Geophysics*, 11(5/6):721–729, 2004.

[97] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of dna nucleotides. *Phys. Rev. E*, 49:1685, 1994.

[98] P. Pišoft, J. Mikšovskỳ, and M. Žák. An analysis of the spatial distribution of approximate 8 years periodicity in ncep/ncar and era-40 temperature fields. *The European Physical Journal Special Topics*, 174(1):147–155, 2009.

[99] G. Plaut, M. Ghil, and R. Vautard. Interannual and interdecadal variability in 335 years of central england temperatures. *Science*, 268(5211):710–713, 1995.

[100] B. Podobnik and H. E. Stanley. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Physical review letters*, 100(8):084102, 2008.

[101] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Communications in Mathematical Physics*, 74(2):189–197, Jun 1980.

[102] A. Rebenshtok, S. Denisov, P. Hänggi, and E. Barkai. Non-normalizable densities in strong anomalous diffusion: Beyond the central limit theorem. *Physical review letters*, 112(11):110601, 2014.

[103] L. F. Richardson. Atmospheric diffusion shown on a distance-neighbour graph. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 110(756):709–737, 1926.

[104] D. Rybski, A. Bunde, and H. von Storch. Long-term memory in 1000-year simulated temperature records. *J. Geophys. Res.*, 113:D02106, 2008.

[105] K. Rypdal. Global temperature response to radiative forcing: Solar cycle versus volcanic eruptions. *Journal of Geophysical Research: Atmospheres*, 117(D6), 2012.

[106] K. Rypdal. Global warming projections derived from an observation-based minimal model. *Earth Syst. Dynam.*, 7:51, 2016.

[107] K. Rypdal, L. Østvand, and M. Rypdal. Long-range memory in earth's surface temperature on time scales from months to centuries. *J. Geophys. Res.*, 118:7046, 2013.

[108] M. Rypdal and K. Rypdal. Long-memory effects in linear response models of earth's temperature and implications for future global warming. *J. Cli.*, 27:5240, 2014.

[109] J. M. Sancho, A. Lacasta, K. Lindenberg, I. M. Sokolov, and A. Romero. Diffusion on a solid surface: Anomalous is normal. *Physical review letters*, 92(25):250601, 2004.

[110] M. Sato, J. E. Hansen, M. P. Mccormick, and J. B. Pollack. Stratospheric aerosol optical depths, 1850-1990. *J. Geophys. Res.*, 98(D12):22,987–22,994, 1993.

[111] B. Schäfer, M. Timme, and D. Witthaut. Isolating the impact of trading on grid frequency fluctuations. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–5. IEEE, 2018.

[112] J. H. P. Schulz, E. Barkai, and R. Metzler. Aging renewal theory and application to random walks. *Phys. Rev. X*, 4:011028, 2014.

[113] S. E. Schwartz. Heat capacity, time constant, and sensitivity of earth's climate system. *J. Geophys. Res.*, 112:D24S05, 2007.

[114] L. Seemann, J.-C. Hua, J. L. McCauley, and G. H. Gunaratne. Ensemble vs. time averages in financial time series analysis. *Physica A: Statistical Mechanics and its Applications*, 391(23):6024 – 6032, 2012.

[115] A. K. Sen and D. Ogrin. Analysis of monthly, winter, and annual temperatures in zagreb, croatia, from 1864 to 2010: the 7.7-year cycle and the north atlantic oscillation. *Theoretical and applied climatology*, 123(3-4):733–739, 2016.

[116] E. P. R. Service. Understanding electricity markets in the eu. 2016.

[117] M. F. Shlesinger and J. Klafter. Lévy walks versus lévy flights. In *On growth and form*, pages 279–283. Springer, 1986.

[118] K. Stein, A. Timmermann, and N. Schneider. Phase synchronization of the el niño-southern oscillation with the annual cycle. *Phys. Rev. Lett.*, 107:128501, Sep 2011.

[119] B. A. Steinman, M. E. Mann, and S. K. Miller. Atlantic and pacific multidecadal oscillations and northern hemisphere temperatures. *Science*, 347(6225):988–991, 2015.

[120] A. K. Tank, J. Wijngaard, G. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, et al. Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment. *International journal of climatology*, 22(12):1441–1453, 2002.

[121] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3(04):785–798, 1995.

[122] M. Thaler. Transformations on [0, 1] with infinite invariant measures. *Israel Journal of Mathematics*, 46(1-2):67–96, 1983.

[123] M. Thaler. Asymptotic distributions and large deviations for iterated maps with an indifferent fixed point. *Stoch. Dyn.*, 5:425, 2005.

[124] D. W. Thompson, J. M. Wallace, P. D. Jones, and J. J. Kennedy. Identifying signatures of natural climate variability in time series of global-mean surface temperature: Methodology and insights. *Journal of Climate*, 22(22):6120–6141, 2009.

[125] E. J. M. Van Den Besselaar, A. M. G. Klein Tank, G. Van Der Schrier, M. S. Abass, O. Baddour, A. F. Van Engelen, A. Freire, P. Hechler, B. I. Laksono, Iqbal, R. Jilderda, A. K. Foamouhoue, A. Kattenberg, R. Leander, R. M. Güingla, A. S. Mhanda, J. J. Nieto, Sunaryo, A. Suwondo, Y. S. Swarinoto, and G. Verver. International climate assessment & dataset: Climate services across borders. *Bulletin of the American Meteorological Society*, 96(1):16–21, 2015.

[126] C. A. Varotsos, M. N. Efstathiou, and A. P. Cracknell. On the scaling effect in global surface air temperature anomalies. *Atmos. Chem. Phys.*, 13:5243–5253, 2013.

[127] K. Y. Wan and R. E. Goldstein. Rhythmicity, recurrence, and recovery of flagellar beating. *Phys. Rev. Lett.*, 113:238103, Dec 2014.

[128] C. Wang, C. Deser, J.-Y. Yu, P. DiNezio, and A. Clement. El niño and southern oscillation (enso): a review. In *Coral Reefs of the Eastern Tropical Pacific*, pages 85–106. Springer, 2017.

[129] T. Weißbach and E. Welfonder. High frequency deviations within the european power system–origins and proposals for improvement. *VGB powertech*, 89(6):26, 2009.

[130] R. Weron. Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and its Applications*, 312(1):285 – 299, 2002.

[131] B. Whitcher, P. Guttorp, and D. B. Percival. Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research: Atmospheres*, 105(D11):14941–14962, 2000.

[132] K. Willson and D. P. Francis. A direct analytical demonstration of the essential equivalence of detrended fluctuation analysis and spectral analysis of rr interval variability. *Physiological measurement*, 24(1):N1, 2002.

[133] K. Yamasaki, A. Gozolchiani, and S. Havlin. Climate networks around the globe are significantly affected by el niño. *Phys. Rev. Lett.*, 100:228501, Jun 2008.

[134] Q. Zhang, Y. Zhou, and V. P. Singh. Detrending methods for fluctuation analysis in hydrology: amendments and comparisons of methodologies. *Hydrological Processes*, 28(3):753–763, 2014.

List of publications

[A] Jeremy O. Richardson, Philipp Meyer, Marc-Oliver Pleinert, and Michael Thoss An analysis of nonadiabatic ring-polymer molecular dynamics and its application to vibronic spectra. *Chemical Physics*, 482:124–134, Jan 2017.

[B] Philipp Meyer and Holger Kantz. Infinite invariant densities due to intermittency in a nonlinear oscillator. *Phys. Rev. E*, 96:022217, Aug 2017.

[C] Philipp Meyer, Eli Barkai, and Holger Kantz. Scale-invariant green-kubo relation for time-averaged diffusivity. *Phys. Rev. E*, 96:062122, Dec 2017.

[D] Philipp Meyer, Marc Hoell, and Holger Kantz. Reproducing long-range correlations in global mean temperatures in simple energy balance models. *Journal of Geophysical Research: Atmospheres*, 123(9):4413–4422, Apr 2018.

[E] Philipp G Meyer, Vidushi Adlakha, Holger Kantz, and Kevin E Bassler. Anomalous diffusion and the moses effect in an aging deterministic model. *New Journal of Physics*, 20(11):113033, Nov 2018.

[F] Philipp G Meyer and Holger Kantz. Inferring characteristic timescales from the effect of autoregressive dynamics on detrended fluctuation analysis. *New Journal of Physics*, 21(3):033022, Mar 2019.

[G] Philipp G Meyer and Holger Kantz. A simple decomposition of European temperature variability capturing the variance from days to a decade. *Climate Dynamics*, 53(11):6909, Dec 2019.

Other manuscripts

[H] Philipp G Meyer and Holger Kantz. From long memory to oscillatory modes – The potentials of detrended fluctuation analysis. *Conference paper*, ISBN: 978-84-17970-78-9, September 2019.

[I] Philipp G Meyer, Mehrnaz Anvari, and Holger Kantz. Identifying characteristic timescales in power grid frequency fluctuations with DFA. *Submitted to Chaos*.

[J] Vidushi Adlakha, Erez Aghion, Kevin E Bassler, Holger Kantz and Philipp G Meyer. Moses, Noah and Joseph Effects and Infinite Densities in Coupled Lévy Processes. *In Preparation*.

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Die Arbeit wurde am Max-Planck-Institut für Physik komplexer Systeme in der Abteilung „Nichtlineare Dynamik und Zeitreihenanalyse" angefertigt und von Prof. Dr. Holger Kantz betreut. Es haben keine früheren erfolglosen Promotionsverfahren stattgefunden. Ich erkenne die Promotionsordnung der Fakultät Mathematik und Naturwissenschaften der Technischen Universität Dresden vom 23.02.2011 in ihrer letzten Änderung vom 06.03.2019 an.

_____

Philipp Meyer   Dresden, 17. Dezember 2019