WESTFÄLISCHE-WILHELMS-UNIVERSITÄT

MASTER THESIS

# Anticipation of critical transitions in complex systems

*Author:*
Frank EHEBRECHT

*Supervisor:*
Dr. Oliver KAMPS

*A thesis submitted in fulfilment of the requirements for the degree of Master of Science*

*in the*

AG Thiele
WWU - Institut für theoretische Physik

March 20, 2017

# Contents

# List of Abbreviations

| | |
|---|---|
| **csd** | **c**ritical **s**lowing **d**own |
| **Mts** | **M**arkov **t**ime **s**cale |
| **HZw** | **H**aken **Z**wanzig |
| **spb** | **s**upercritical **p**itchfork **b**ifurcation |
| **CD dataset** | **C**hristopher **D**anforth dataset |
| **BV dataset** | **B**onne**V**ille dataset |
| **CKe** | **C**hapman **K**olmogorov **e**quation |
| **acf** | **a**uto **c**orrelation **f**unction |

# Chapter 1

# Introduction

Complex systems in nature may exhibit spontaneous changes of stability from one state into another. At such so-called *tipping points* the dynamical properties change qualitatively. It has been an ongoing concern to find measures to predict such *tipping points*, as the *transition* into another state may carry along unwanted properties.
The general concept is to examine the data of past events or some simulated data, to model the system, find measures to forecast the *transition* or to establish general concepts, to anticipate future events.
Such *early-warning signals* have been searched for in a lot of different dynamical systems, not solely limited to the field of physics. Publications have been made examining power outages in power systems [1], [2], earthquakes from seismic data [3], [4], epileptic seizures [5], climate change [6], [7], [8], [9], economic data [10], ecological systems [11], onset of war [12], medicine [13], [14] and others.
The methods applied in these publications persue different but reoccurring approaches, that may be categorized by their ansatz (approaches). Among these are

- measuring the stochastic properties,
- determining the time scales of a process,
- modelling of dynamics with drift and diffusion estimation,
- examination of the Markov property.

As measures of stochastic properties there have been used simple stochastic moments [1],[5],[15] or pdf-based metrics [3]. A widespread time scale metric is the lag-1 autocorrelation, called the $a_1$-coefficient. It measures the mean linear dependance of consecutive time steps and has been used in [15], [6], [1], [5].
A modelling of drift and diffusion provides insight into the dynamics of the underlying process and therefore gives an eidedic approach of connecting results to bifurcation theory. An application of this has been made in [15],[16],[17]. With help of the drift one can also estimate the potential of a process and use this as measure.
Examining the Markov property means to determine on what time scale a process can be described or described best as a Markov process [4]. Regarding only one variable in some large complex system most certainly leads to a non-Markovian property in such that the best prediction of a time series not only

depends on the latest time step but additional time steps from the past yield a better prediction.
To apply these measures to time series one can use 'sliding time-windows' where one applies the measures on some interval and shifts this interval in time to detect a change. The advantage of the sliding window method is its straightforeward applicability but on the downside, to fully react to a change in parameters it needs the width of one time window and it has to be interpreted what this change means for the parameter. Another method that can be applied in sliding windows but can also be applied to the complete data set is Bayesian analysis [7],[8], where one first has to make additional assumptions about the kind change - e.g. linear trend, jump, etc. Then one can apply the measure and with the help of Bayesian statistics one can determine how well the assumption is fulfilled or at what point the assumption is fulfilled best. If a suitable assumption about the kind of change has been made, Bayesian analysis yields a more accurate determination of the position of that change, but constricts the applicability by the need of just this assumption about the transition. A direct (*online*) usage for *predicting* transitions is unsuitable, since one needs sufficient datapoints *after* the transition.
The informative value of such measures must always weight up with the amount of data points of the sampling interval. The variance for example is just one scalar measure that uses all datapoints to estimate its value and therefore provides good statistics. A drift estimation indeed has higher informative value but does not yield *one* scalar value but a series of coefficients and therefore provides worse statistics on the same sample interval, if one not constructs a scalar measure from the coefficients.
All these methods measure some property of the underlying system, whereby the properties mentioned above may have overlapping implications. This leads to the assumption, that there are general concepts that are believed to hold true for a wide range of transitions and can be seen in all measures. In [18] M. Scheffer proposes different kinds of systems, where 'highly connected' and 'homogeneous' systems show the occurrence of critical transitions. Such systems are resilient to change, since small disturbances get mended by the rest of the system, but a build up of stress leads to a point of spontaneous collapse. A generally accepted symptome of this is the concept of *critical slowing down* [5]. In the context of *critical transitions*, this means that the system 'becomes increasingly slow in recovering from small perturbations' [5] as it approaches the bifurcation point, since the potentials fatten. This behaviour has been shown for many systems.

Although complex systems consist of a high order of dynamical interactions, in real life systems one often has only access to one or a few measurable quantities. This can be for example the voltage frequency for power systems, or the seismograph amplitudes for earthquakes, but those hold important information about the system.
If one observes such a quantitiy over an extended period of time prior to a tipping point, it will have a noisy dynamic with a constant or very slowly

changing mean value. There may be measures or indicators, that can be directly spotted or suspected by only looking at the time series, but there may also be qualities that change although the time series has seemingly constant stoachstic properties. This means that these properties can only be detected after application of the measures in the first place.

In this master thesis the times series of power outage data will be examined with different methods with the aim to understand the mechanics of the outage and to find the measures to predict it. This data will be the bus voltage frequency of power systems before an outage. An advantage of such systems is the relatively high amount of datapoints for one event compared to other systems and evenly spaced time increments that make the analysis easier to handle. As preliminary analysis synthetic data sets will be analyzed, to test the measures.

# Chapter 2

# Theory

## 2.1 Critical transitions in complex systems

The central appeal of this master thesis is to find measures to forecast critical transitions in stochastic complex systems. In general the data is a time series of some quantity that is needed to be retained at a desired value over time – e.g. a constant voltage frequency in power systems.
From a technical or applied viewpoint the observed quantity has to remain at a certain value regulated by some complex control mechanisms. From a mathematical viewpoint the observed quantity then can be interpreted as the time behaviour of one of the dynamical variables $x_1$ of a set of stochastic ordinary differential equations (eq. 2.1), with the requirement of $x_1$ describing a stationary process in some regime, for constant $\boldsymbol{\epsilon}$. Thus one turns a technical necessity into a mathematical principle.

$$\begin{aligned} \dot{x}_1 &= f_1(\mathbf{X}, \boldsymbol{\epsilon}) + \Gamma_1(t) \\ \dot{x}_2 &= f_2(\mathbf{X}, \boldsymbol{\epsilon}) + \Gamma_2(t) \\ \vdots &\quad = \vdots \\ \dot{x}_n &= f_n(\mathbf{X}, \boldsymbol{\epsilon}) + \Gamma_n(t) \end{aligned} \tag{2.1}$$

In general the $f_n$ are some nonlinear functions dependent on the dynamical variables $\mathbf{X} = (x_1, x_2, \ldots, x_n)^T$ and a set of parameters $\boldsymbol{\epsilon}$. Additionally there are random fluctuations $\Gamma_i$.
Assumptive the system runs in normal operation, the set of parameters $\boldsymbol{\epsilon}$ are (nearly) constant over time. If the system now for some (external) reason approaches a system failure, e.g., a power outage in a power system, this can be expressed by a change in $\boldsymbol{\epsilon}$ up to a 'point' where the mode of operation changes qualitatively, i.e., a bifurcation. Since for constant $\boldsymbol{\epsilon}$, the time evolution of $x_1$ describes a stationary process, the crossing of the bifurcation point can be called a phase transition.
If one considers $x_1$ as the only available (measurable) quantity, there arise two kinds of incompleteness of knowledge about the state of the system. That is on the one hand the influence of random fluctuations $\Gamma_i$, and on the other hand the neglection of all other $x_i$. To explicate this notion 'of incompleteness of knowledge' further, one has to treat this problem in a probabilistic manner.
The time evolution of a collection of random variables as in (eq. 2.1) is some complex stochastic process. An important kind of stochastic processes are Markov

processes. These are memoryless ,i.e., the probabilistic future can be predicted solely on its present state. Any additional time steps from the past will not yield any better prediction.
The prime example of a Markov process is Brownian Motion ([19], [20], [21], [22] and others). This is the 'motion of a heavy particle immersed in a fluid of light molecules'[22]. Those light molecules move due to thermal motion and the heavy particle is constantly hit, which leads to a random motion. One now observes the time evolution of the positions $X$ of the heavy particle. On a timescale where one can resolve each of the small collisions one checks if the Markov assumption holds. Given, the heavy particle had a relatively large displacement of $x_k - x_{k-1}$, chances are the velocity at $x_k$ is also relatively large. This velocity will survive for a short time (due to conservation of momentum), thus a larger $x_{k+1} - x_k$ will be favoured. This means on this timescale this stochastic process is not Markovian, since $x_{k+1}$ is not solely correlated to $x_k$, but also to $x_{k-1}$. If one now considers this problem on a coarse-grained time scale, each step of the heavy particle contains many molecule collisions (see figure 2.1 b & c). On this time scale the conservation of momentum can be neglected, thus the process can be assumed to be Markovian again. In this example we only looked at the position of the heavy particle. If one also includes the positions and velocities of all small particles, one would of course regain the Markov property (this corresponds to taking along all $x_i$ from (eq. 2.1)) on all time scales.

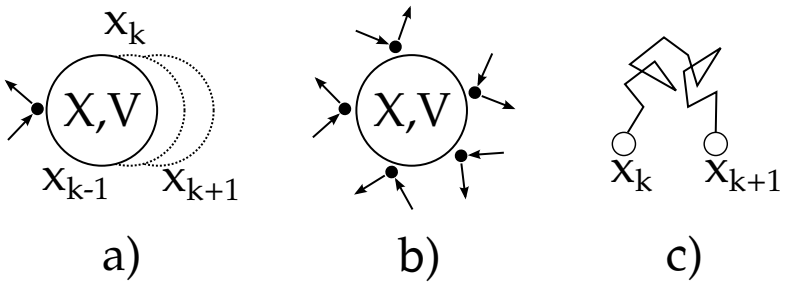


FIGURE 2.1: Brownian motion. a) three time sequentially positions of the heavy particle on a very finely resolved time scale. b) & c) the same process as in a), but after many collisions.

However, a system as the Brownian motion system consists of too many parts as to solve all these 'microscopic' equations (as one only has $x_1$ as measured quantity).
To fully describe such a system one would have to solve a plethora of equations. This of course would describe the whole system deterministically on a microscopic level. Because one generally neither has the information about the full set of microscopic equations nor the mathematical methodology to handle them, one uses some macroscopic variables which fluctuate in a stochastic manner, thus describing the system stochastically. This way one describes the time evolution of some probability distribution. Transitioning from probability densities to some expected values, one can describe the system deterministically on a macroscopic level (see fig. 2.2 taken from [19]).
The objective is now to analyze the time series of some measured quantity of

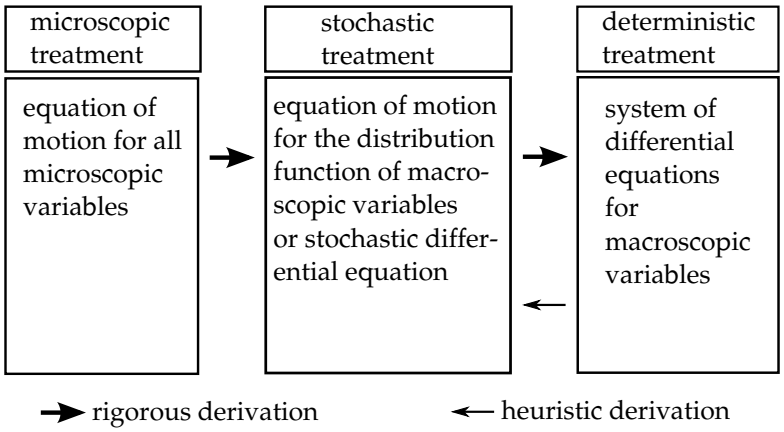


FIGURE 2.2: Classification as to treat complex systems (taken from [19])

a complex system, and to find measures that indicate that a transition is imminent. To illustrate this with a simplified example, we take a supercritical

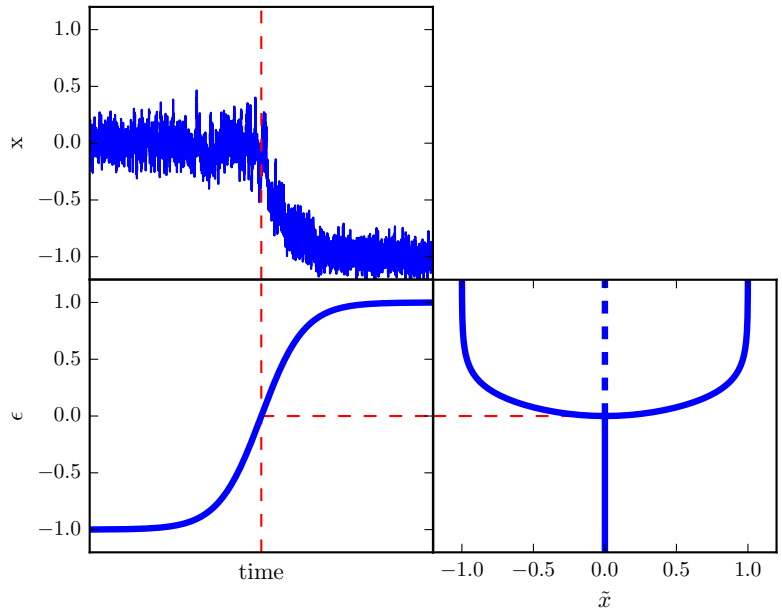


FIGURE 2.3: Supercritical pitchfork with additive noise. top left: sample path of this bifurcation; bottom left: change of the bifurcation parameter; bottom right: bifurcation diagram for given control parameter.

pitchfork bifurcation with additive noise

$$\dot{x} = \epsilon x - x^3 + q\,\Gamma(t), \tag{2.2}$$

and simulate such a process with varying bifurcation parameter $\epsilon$ (bottem left of figure 2.3). A sample path can be seen in the top left plot. The point of bifurcation is marked with dotted lines. One now wants to examine the stochastic properties of such a sample path before the bifurcation to anticipate its imminence.

### 2.1.1 Slaving principle

The slaving principle is a term stemming from self-organization. As a starting point we take equation 2.1 and neglect the fluctuating parts. In Haken [23] it is shown, that one can transform such an system of equations with the help of linearisation into

$$\begin{array}{rl} \dot{x}_1 & = -\gamma_1 x_1 + \tilde{f}_1(\mathbf{X}, \tilde{\boldsymbol{\epsilon}}) \\ \dot{x}_2 & = -\gamma_2 x_2 + \tilde{f}_2(\mathbf{X}, \tilde{\boldsymbol{\epsilon}}) \\ \vdots & \quad = \vdots \\ \dot{x}_n & = -\gamma_n x_n + \tilde{f}_n(\mathbf{X}, \tilde{\boldsymbol{\epsilon}}) \end{array} , \tag{2.3}$$

with $\tilde{f}_n$ being nonlinear functions of $\mathbf{X}$, with explicitly no linear part. The variables can now be split into two groups: those with small damping that can even become unstable ($\gamma_i \lesssim 0$), and the other group being the stable modes with $\gamma_s \gg 0$. This partition is not uniquely defined. With these properties we can use the adiabatic approximation principle ([23]), setting $\dot{x}_s = 0$. We also assume, that $|x_s| \ll |x_i|$. With this we can set $x_s = 0$ in all $\tilde{f}_s$, eliminating all $\dot{x}_s = 0$ in equations 2.3. This yields

$$\dot{x}_i = -\gamma_i x_i + \tilde{f}_i \left[\mathbf{X}_i; \mathbf{X}_s(x_i)\right] . \tag{2.4}$$

In this way only the unstable modes remain to describe the system. In Haken [23] it says that, 'in many practical applications only one or very few modes become unstable'. This means that the complex system can largely be described by only one or a few order parameters. Taking the fluctuations back into the equation, exactly this property makes such a system interesting for stochastical analysis.

### 2.1.2 Critical slowing down

Critical slowing down (CSD) is a term stemming from non-stochastic systems and describes the 'lethargic decay'[24] at the bifurcation point of, for example, a pitchfork bifurcation. This concept can also be applied to stochastic systems. If one takes a system with a restoring force (the deterministic part) and some fluctuating force (small random perturbations), the phenomenon exhibited is a wider range of the dynamics in phase space, since the deterministic force doesn't restore the dynamic as strong near a transition as it would do far from it, yet the fluctuating force consistently drives the systems (see fig. 2.4). This will be further discussed in chapter 4.

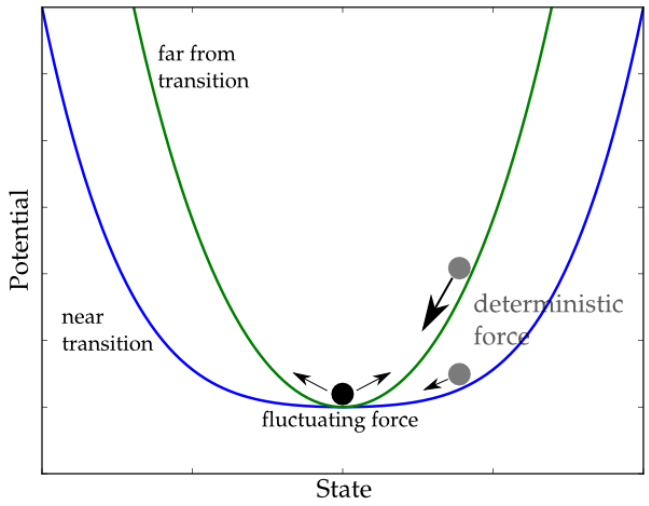


FIGURE 2.4: Critical slowing down (as in [5]). The potential of the dynamical system gets flatter as the bifurcation is being approached. This leads to a spread of the dynamics in phase space.

## 2.2 Stochastic properties

### 2.2.1 Random variables and probability measures

The definition of a random variable as given in [25] p.184 is:

- A function $X$, that has every elementary event $\omega \in \Omega$ a real number $x$ uniquely assigned to it, so that for every $x \in \mathbb{R}$ the probability $\Pr(X(\omega) \leq x)$ is defined, is called a **random variable**.

Such a random variable can be discrete or continuous. A discrete random variable has a finite or a countable infinite amount of realisations $x_i$ and can be represented by its probability function

$$Pr(X = x) = \begin{cases} Pr(X = x_i) =: P_i & \text{for } x = x_i,\ i = 1, 2, ... \\ 0 & \text{for } x \text{ else} \end{cases} . \tag{2.5}$$

A continous random variable has an uncountable set of realisation with a density function $f(x)$ that fulfills

- $-\infty < x < +\infty$
- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1.$

A random variable $X$ is always stochastically completely known, if one knows its distribution.

### 2.2.2 Law of large numbers & central limit theorem

Limit theorems appear in the

- '**laws of large numbers**: Statements can be made about the properties of empirical distributions and their parameters as the sample size increases.'[25]
- '**central limit theorems**: Statements can be made under which conditions a series of distributions converges to a normal distribution.'[25]

There are several different laws of large numbers and central limit theorems but the most important properties are

- that the distribution of some measured stochastical quantity can be the better determined the more data points one uses and
- that the sum of n identical independent random variables with the same standard variation $\sigma$ and expexted value $\mu$ convergeres towards the normal distribution
$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-0.5\left(\frac{x-\mu}{\sigma}\right)^2} \tag{2.6}$$
as $n \to \infty$.

This is the most common property of the central limit theorems, but there are several additional theorems to expand the conditions that lead to a normal distribution.
The random variable of a normal distribution (2.6) can always be rescaled to a random variable with mean zero and variance one. Let $X \sim \mathcal{N}(\mu, \sigma^2)$, than one can get a random variable $Z \sim \mathcal{N}(0, 1)$ by

$$Z = \frac{X - \mu}{\sigma}. \tag{2.7}$$

### 2.2.3 Stochastic moments

Let $X$ be a random variable and $k$ a natural number, then the moment $m_k$ of order $k$ of $X$ is

$$m_k := \mathrm{E}(X^k). \tag{2.8}$$

These moments give the possibility to describe a random variable parameterwise, and to compare random variables of different distributions. In the case of a continuous distribution with density $f_X$ the moments are given as

$$m_k := \int_{-\infty}^{\infty} x^k\, f_X(x)\mathrm{d}x. \tag{2.9}$$

Besides, there are the *central* moments. These are the moments about the expected value of the random variable

$$\mu_k := \int_{-\infty}^{\infty} (x - m_1)^k\, f_X(x)\mathrm{d}x. \tag{2.10}$$

These central moments are the measures of interest if one wants to describe certain characteristics of a random variable, because they offer a very eidetic characterization. The first moment $m_1$ is the expected value, and $\mu_2$ is the variance that measures the width or spread of some distribution. The third central moment $\mu_3$ is called the skewness and measures the asymmetry of a distribution (a value equal zero corresponds to a symmetric distribution and a value greater zero to a right-skewed distribution). The fourth central moment $\mu_4$ is called the kurtosis and measures the arching of a distribution (a value greater zero corresponds to a steep distribution and a value less than zero corresponds to a flat distribution).

### 2.2.4 Covariance & autocorrelation

The covariance is a measure to determine the monotonous correlation between two random variables ($X$ and $Y$) and is defined as

$$\mathrm{Cov}(X,Y) := \mathrm{E}\left[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right]. \tag{2.11}$$

Its value is positive when $X$ and $Y$ have a monotonous correlation, that is big values of $X$ go along with big values of $Y$. The covariance is negative, when big values of $X$ go along with small values of $Y$ and vice versa.
The covariance can be regarded as the generalization of the variance as

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X). \tag{2.12}$$

Since the covariance is a comparative measure that can take arbitrary value depending ond $X$ and $Y$, a standardized measure to make absolute statements of correlation is of interest. This can be implemented by the standard deviations of $X$ and $Y$. This is called the correlation coefficient

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \tag{2.13}$$

with

$$\rho \in [-1,1], \tag{2.14}$$

and its estimator is

$$\hat{\rho}_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}. \tag{2.15}$$

With the help of the correlation coefficient one can define yet another measure, that quantifies how much a discrete time dependant variable resembles itself under a shift of time. This is called the autocorrelation. Given some time dependant variable $x_t$, the autocorrelation of lag $l$ is defined as

$$\hat{\rho}_l = \frac{\sum_{t=1}^{n-l}(x_t - \bar{x})(x_{t+l} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}. \tag{2.16}$$

### 2.2.5 Information & entropy

The notion of information can be interpreted as a physical quantity, as the notion of entropy can be interpreted as a quantity in information theory.
To elaborate these ideas, we consider the result of a random experiment and assign an amount of information to it, interpreting the outcome as the receipt of a message. If we now look at two independent random variables (messages) with discrete uniform distributions with $R_1$ and $R_2$ outcomes respectivly, there are $R_1 \cdot R_2 = R_0$ equally probable outcomes if we look at the joint execution of both experiments. Since 'information' (I) must be based on the amount of possibilities of outcomes one has to demand

$$I(R_1 \cdot R_2) = I(R_1) + I(R_2). \tag{2.17}$$

This of course is fulfilled by

$$I(R_0) = K \cdot \log(R_0). \tag{2.18}$$

The factor K can be set by choosing a logarithm with the base $a$, being the amount of possibile outcomes. If a sign x has the probability of occurrence $p_x$, then

$$I(x) = -\log_a(p_x). \tag{2.19}$$

This quantity is called self-information and ties common events to low self-information and rare events to high self-information. If one wants to get the *mean* self-information of a message, one has to weigh the self-information with its probability of occurrence and sum over all events. This is called entropy

$$H = -\sum_{x \in X} p_x \log_a(p_x). \tag{2.20}$$

## 2.3 Stochastic processes

The definition of a stochastic process as in [20] is:

- A stochastic process is a family $(X_t)_{t \in I}$ of random variables.

The index set $I$ can be discrete, leading to a discrete-time stochastic process or it can be continuous, leading to a continuous-time stochastic process.

### 2.3.1 White noise process

A white noise process is a normally distributed stochastic process with zero mean and without any correlation between its elements. This means every time step is independent from all others. Some of its basic properties are

- The fourier transform of a white noise process, again is a white noise process.
- The autocorrelation of a white noise process is a Dirac-delta function.

### 2.3.2 Wiener process

A very important type of 'stochastic process' is the Wiener process. It is essential for the description of stoachstic differential equations. A derivation and definition as in [26] will be used here. First we define some equidistant, disjunct partition of an index set

$$[0,1) = \bigcup_{i=1}^{n} \left[ \frac{i-1}{n}, \frac{i}{n} \right). \tag{2.21}$$

We now define a random walk with $\epsilon_t \in \{-1, 1\}$

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} \epsilon_j \;\; \text{for}\, t \in \left[ \frac{i-1}{n}, \frac{i}{n} \right), i = 1, ..., n. \tag{2.22}$$

This is a step function with increments $\pm\frac{1}{\sqrt{n}}$ and step width $\frac{1}{n}$ on the interval $[0, 1)$. The limit $n \to \infty$ yields the Wiener process:

$$X_n(t) \Rightarrow W(t) \;\; \text{for}\; n \to \infty. \tag{2.23}$$

This limit contains a special concept of convergence ([26], chapter 13) and leads to a process continous in space and time (but not differentiable). This Wiener process $W(t)$, $t \in [0, T]$ fulfills

- The starting value is zero with a probability of one, $P(W(0) = 0) = 1$.
- Non-overlapping increments $W(t_1) - W(t_0), ..., W(t_n) - W(t_{n-1})$, are independant for arbitrary $n$.
- The increments are normal distributed with $W(t) - W(s) \sim \mathcal{N}(0, t - s)$, with $0 \leq s < t$.

### 2.3.3 Stationary processes

A process $X_t$ is called weakly stationary if

- the expected value is constant $\mathrm{E}(x_t) = \mathrm{const}$ for all $t$
- the variance is finite for all $t$
- the covariance is independent of a shift in time ($\mathrm{Cov}(x_{t_1}, x_{t_2}) = \mathrm{Cov}(x_{t_1+\Delta t}, x_{t_1+\Delta t})$).

A process is called strongly stationary, if its probability distribution is independent of any temporal shifts.

### 2.3.4 Markov property & Markov processes

The Markov property was already addressed in chapter 2 and will be elaborated here. Given some stochastic process $\mathbf{X}_t$ with

$$t_n \geq t_{n-1} \geq \cdots \geq t_1 \tag{2.24}$$

the transition probability

$$p(\mathbf{x}_n, t_n \,|\, \mathbf{x}_{n-1}, t_{n-1};\, \ldots ; \mathbf{x}_1, t_1) \tag{2.25}$$

describes the process completely. The Markov property now demands that any additional time steps from the past except the lastest step will not yield any additional information. For the transition probability this means:

$$p(\mathbf{x}_n, t_n \,|\, \mathbf{x}_{n-1}, t_{n-1};\, \ldots ; \mathbf{x}_1, t_1) = p(\mathbf{x}_n, t_n;\, |\, \mathbf{x}_{n-1}, t_{n-1}). \tag{2.26}$$

While the index set here was taken as discrete, one can also assume time-continuous Markov processes. For a variable as $x_1$ in 2.1 the Markov property would almost certainly not hold on arbitrarily small time steps, since 'equations which are derived are rarely truly Markovian - rather there is a certain characteristic memory time during which the previous history is important'[21]. This means that there is really no such thing as a Markov process; rather, there may be systems whose memory time is so small that, on the time scale on which we carry out observations, it is fair to regard them as being well approximated by a Markov process'([21], chapter 3.3). As mathematical definition one can demand for a continuous Markov process with continuous sample paths

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\mathbf{x}-\mathbf{z}|>0} p(\mathbf{x}, t+\Delta t|\mathbf{z}, t) d\mathbf{x} = 0, \tag{2.27}$$

for $\epsilon > 0$.

### 2.3.5 Stochastic differential equations & Langevin equations

With the proceeding definitions one can formulate the concept of stochastic differential equations

$$\mathrm{d}X_t = a(X_t, t)\,\mathrm{d}t + b(X_t, t)\,\mathrm{d}W_t. \tag{2.28}$$

This differential equation describes a stochastic process $X_t$, that resembles an ordinary differential equation in the $\mathrm{d}t$-part and is non-deterministic in the $\mathrm{d}W_t$-part, where $W_t$ is a Wiener process. Note, that difference quotient was used, since stochastic processes are non-differentiable. As the Wiener process is used in stochastic differential equations an equivalent definition often applied in physics is the Langevin equation, using a 'rapidly fluctuating random term' $\Gamma(t)$([21]).

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(x, t) + b(x, t)\,\Gamma(t). \tag{2.29}$$

For this random term one demands

$$\langle \Gamma(t) \rangle = 0 \tag{2.30}$$

and

$$\langle \Gamma(t)\Gamma(t') \rangle = \delta(t - t'). \tag{2.31}$$

This definition neglects the fact, that the stochastic process (although continuous) is not differentiable. ([21])

### 2.3.6 Chapman-Kolmogorov equation for Markov processes

Any Markov process as in chapter 2.3.5 can be expressed by its transition probability (eq. 2.26). In return a necessary condition for a Markov process is

$$p\left(\mathbf{x}_1, t_1 | \, \mathbf{x}_3, t_3\right) = \int p\left(\mathbf{x}_1, t_1 | \, \mathbf{x}_2, t_2\right) \, p\left(\mathbf{x}_2, t_2 | \, \mathbf{x}_3, t_3\right) \, \mathrm{d}\mathbf{x}_2. \tag{2.32}$$

This is called the Chapman-Kolmogorov equation. It must be fulfilled for all $t_3 < t_2 < t_1$, for $\mathbf{x}_t$ being a Markov process.

### 2.3.7 Fokker-Planck equation & Kramers Moyal expansion

A Langevin equation describes a set of sample paths of a stochastic process. Another possibility to describe this would be the description of the time evolution of a probability density function $W$ using some unknown transition probability $P$

$$W(x, t+\tau) = \int P(x, t+\tau | x', t) W(x', t) \mathrm{d}t. \tag{2.33}$$

To accomplish this one can expand the distribution function and form the limit $\tau \to 0$ ([19])

$$\frac{\partial W(x,t)}{\partial t} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^n D^{(n)}(x,t) W(x,t) \tag{2.34}$$

where

$$D^{(n)}(x,t) = \lim_{\tau \to 0} \frac{1}{n!\tau} \int (x - x')^n P(x, t+\tau | x', t) \mathrm{d}x \tag{2.35}$$

are the Kramers Moyal coefficients. It can be shown, that for Langevin equation of the form 2.28 the expansion stops after the second term (Pawula theorem [19]). This is then called the Fokker-Planck equation

$$\dot{W}(x,t) = \left(-\frac{\partial}{\partial x} D^{(1)}(x,t) + \frac{\partial^2}{\partial x^2} D^{(2)}(x,t)\right) W(x,t). \tag{2.36}$$

Its connection to the Langevin equation is

$$D^{(1)}(x,t) = a(x,t) + \frac{\partial b(x,t)}{\partial x} b(x,t), \tag{2.37}$$

$$D^{(2)}(x,t) = b^2(x,t), \tag{2.38}$$

$$D^{(n)}(x,t) = 0 \;\; \text{for } n \geq 3. \tag{2.39}$$

Equation 2.35 can be expressed with the help of a local average, since the transition distribution $P$ weights the increments $(x-x')^n$ with their relative frequency of occurrence:

$$D^{(n)}(x,t) = \frac{1}{n!}\lim_{\tau\to 0}\frac{1}{\tau}\left\langle [x(t+\tau)-x(t)]^n \right\rangle . \tag{2.40}$$

### 2.3.8 Ornstein-Uhlenbeck process

A process with linear drift and constant diffusion

$$dX_t = -\theta\, X_t\, dt + \sigma\, dW_t \tag{2.41}$$

is called Ornstein-Uhlenbeck process. If one takes the starting value $X_0 = a$, the expected value becomes

$$E(X_t) = ae^{-\theta t}, \tag{2.42}$$

and the covariance reads as

$$\gamma := \mathrm{Cov}(X_s, X_t) = \frac{\sigma^2}{2\theta}\left(e^{-\theta|s-t|} - e^{-\theta(s+t)}\right). \tag{2.43}$$

Thus with $\mathrm{Var}(X_t) = \mathrm{Cov}(X_t, X_t)$ the variable $X_t$ is distributed as

$$X_t \sim \mathcal{N}\left(ae^{-\theta t}, \frac{\sigma^2}{2\theta}(1-e^{-2\theta t})\right). \tag{2.44}$$

Since the expected value and the variance converge, there exists a stationary distribution:

$$X_{t,stationary} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{2\theta}\right). \tag{2.45}$$

Furthermore the autocorrelation for an Ornstein-Uhlenbeck process can be formulated with $\Delta t = t - s$ with $t, s \to \infty$ and $s < t$

$$\rho_{\text{stationary}} = \lim_{s\to\infty}\frac{\gamma_{\Delta t}}{\gamma_0} = \lim_{s\to\infty}\frac{e^{-\theta\Delta t} - e^{-\theta(\Delta t+2s)}}{1-e^{-2\theta s}} = e^{-\theta\Delta t}. \tag{2.46}$$

If one assumes a time discrete time series with a sampling interval of $\Delta t_{samp}$, one can estimate for what combinations of $\theta$ and $\Delta t_{samp}$, a reconstruction of the process via the autocorrelation is still possible. Under the assumption, that the limit of reconstruction is when after the first step, the value has declined to $\frac{1}{e}$, we get

$$e^{-\theta\, 0} = \frac{1}{e}e^{-\theta\Delta t_{samp}}, \tag{2.47}$$

$$\theta \leq \frac{1}{\Delta t_{samp}}. \tag{2.48}$$

This is depicted in figure 2.5.

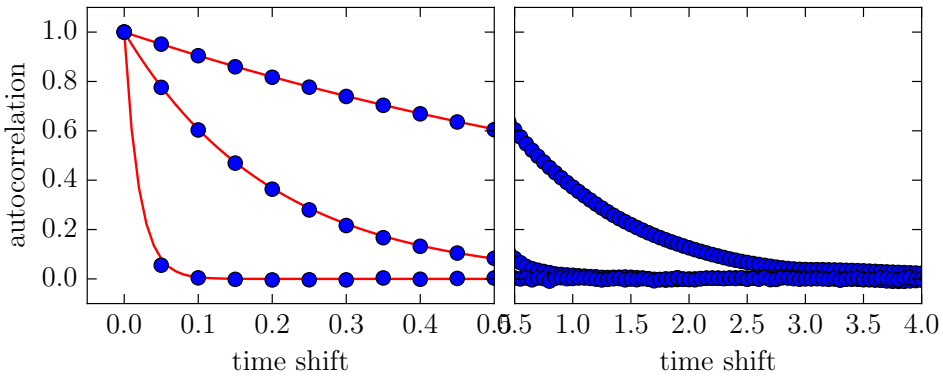


FIGURE 2.5: The autocorrelation for three Ornstein Uhlenbeck processes, with $q = 1$ and $\gamma = 1, 5, 50$

## 2.4 Analytical methods

### 2.4.1 Euler-Maruyama method

To calculate sample paths of stochastic processes one uses what could be called the stochastic analog to the simple Euler method for ODEs. If one has some stochastic differential equation of the form

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dW_t \tag{2.49}$$

one can apply the Euler step to the deterministic part. The stochastic part can be determined by third property of the Wiener process and the transformation rule for normal distributions 2.7:

$$W_{t+\Delta t} - W_t \sim \mathcal{N}(0, \Delta t) \sim \sqrt{\Delta t}\, \mathcal{N}(0, 1). \tag{2.50}$$

Altogether:

$$X_{t+\Delta t} = X_t + a(t, X_t)\, \Delta t \; + \; b(t, X_t)\, \mathcal{N}(0, 1)\, \sqrt{\Delta t}. \tag{2.51}$$

This method is suitable for all calculations. Although there are more complex methods that are more precise per time step, these methods on the downside are by far more complicated. This method converges strongly towards the searched sample path. This means one can always chose a $\Delta t$ small enough, to calculate a sufficiently accurate sample path.

### 2.4.2 Kernel regression & Kernel density estimation

In this section several non-parametric methods to estimate functional relation of data are presented. As base we start with the kernel density estimation, that

delivers an estimation of the pdf $\hat{f}$ of a given data set $x_i$

$$\hat{f}(x) = \frac{1}{N\,\sigma} \sum_{i=1}^{N} \mathrm{K}\left(\frac{x - x_i}{\sigma}\right), \tag{2.52}$$

with $N$ datapoints and some kernel $\mathrm{K}$ with width $\sigma$. The kernel can have several forms like Gaussian or Epanechnikov

$$K_{epan}(y) = \begin{cases} \frac{3\sqrt{5}}{100}(5 - y^2) & \text{if } y^2 < 5 \\ 0 & \text{if } y^2 > 5 \end{cases}. \tag{2.53}$$

A common estimation of the width is Silverman's rule of thumb

$$\sigma = \left(\frac{4\hat{\sigma}^5}{3N}\right)^{\frac{1}{5}}, \tag{2.54}$$

with $\hat{\sigma}$ being the standard deviation of the dataset. A closely related method, the Nadaraya Watson regression, to get a (non-parametric) regression function $\hat{y}$, weights the local function mean of a data set $(x_i|y_i)$ with a kernel

$$\hat{y}(x) = \frac{\sum_{i=1}^{N} y_i\, \mathrm{K}(\frac{x-x_i}{\sigma})}{\sum_{i=1}^{N} \mathrm{K}(\frac{x-x_i}{\sigma})}. \tag{2.55}$$

This kind of regression can also be used to smooth a noisy data set to only maintain the big scale trends (called Kernel smoothing). If one subtracts the smoothed data $\hat{y}_{GKS}$ from the original data $y$, one gets a detrended time series with the average set to zero (Gaussian kernel filter - GKF)

$$y_{GKF} = y - y_{GKS}. \tag{2.56}$$

This removes big scale trends, and leaves the increments intact.

# Chapter 3

# Methods of analysis

The methods used in this thesis get applied to some time series in *sliding time windows*. This means one evaluates the measures on all values on the interval $[t_{start}, t_{start} + t_{window}]$ and examines how they change as the window is shifted foreward in time $[t_{start} + t_{shift}, t_{start} + t_{window} + t_{shift}]$. All measures use the assumption, that the process is stationary within the time window. This is of course an approximation, considering one explicitly searches for some stochastical change over time. Thus one has to weight up stochastical accuracy against the sensitivity, as depicted in figure 3.1. Here one has some Gaussian white noise with variable standard deviation in the upper part. In the lower part the standard deviation is depicted as dashed line. One now evaluates the sampling standard deviation for a time window of 100 steps and 1000 steps respectively (black vertical lines). One can clearly see, that $\sigma$ for a small time window reacts faster to a change of stochastical properties, but on the downside is more volatile.
As differently sized time windows effect the accuracy and sensitivity of a measure, the form of different irregularities in parameter determines (gives) the functional form of some window-averaged measure. This is shown and commented in figure 3.2. Strictly, this only works if the measured quantity enters linearly with the window size, as it is for example the case with the variance. If one has some timeseries with two sections of different variance (see figure 3.3) with ${}^aN$ datapoints of ${}^ax_i$ and ${}^bN$ datapoints of ${}^bx_i$, then for the variance over both sets one can say

$$\mathrm{Var}(X) = \frac{1}{N}\sum_i x_i^2 = \frac{{}^aN}{{}^aNN}\sum_i {}^ax_i^2 + \frac{{}^bN}{{}^bNN}\sum_i {}^bx_i^2 = \frac{{}^aN}{N}\mathrm{Var}({}^aX) + \frac{{}^bN}{N}\mathrm{Var}({}^bX). \tag{3.1}$$

Thus the variance enters linearly with the amount of data points. While this is valid for the variance, it does not hold for the Kullback-Leibler divergence. Since $p$ is the same for all windows, we only have to check the second part and since the union of the pdfs is linear

$$\frac{{}^aN}{N}q_{{}^aX} + \frac{{}^bN}{N}q_{{}^bX} = q_X, \tag{3.2}$$

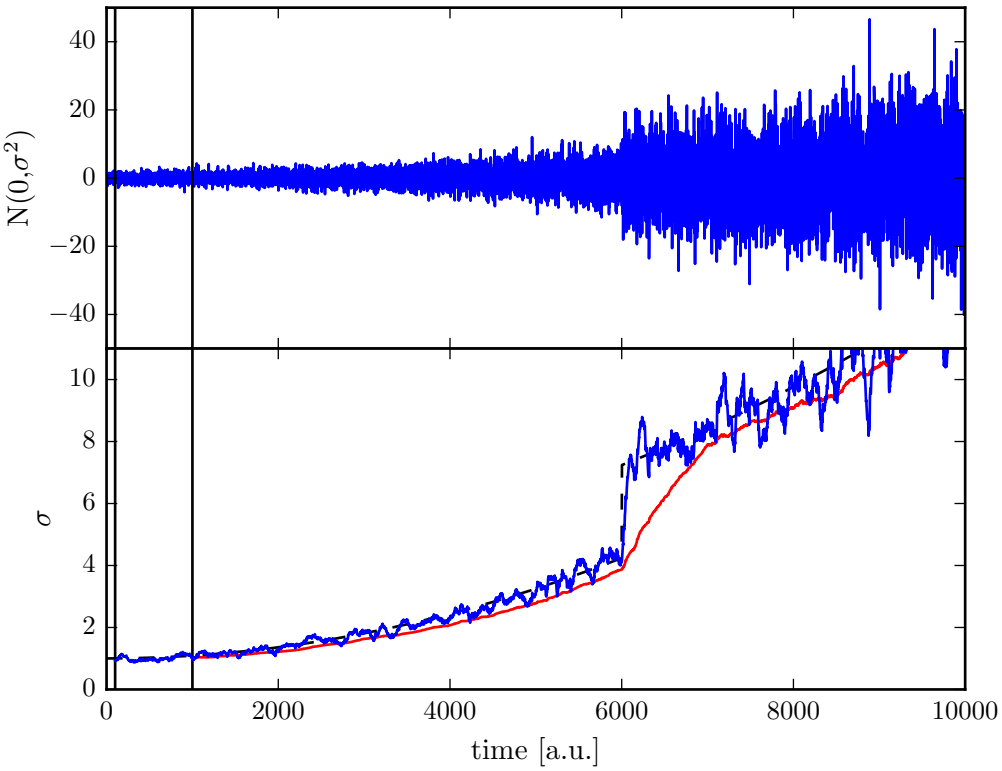


FIGURE 3.1: In the upper part, one sees uncorrelated Gaussian white noise with variable standard deviation. The standard deviation to produce this noise is depicted as dashed line in the lower part, alongside the values for the calculated standard deviation for sliding windows of size 100 steps and 1000 steps respectively.

one only has to see that

$$\frac{^aN}{N}\sum {}^ap\,\log({}^aq)+\frac{^bN}{N}\sum {}^bp\,\log({}^bq)\neq\sum p\,\log(q). \tag{3.3}$$

The methods in this chapter will not be applied to the time series itself, but to some augmented versions of the time series. Firstly, these are time series detrended as described in eq. 2.56, moving the local average to zero and removing big scale trends.
Secondly, these are the increment time series

$$x_{incr,n}=x_{n+1}-x_n, \tag{3.4}$$

taking the increments of each step as datapoints. This also sets the average to zero and removes trends, but foremost it contains another information value as the time series itself. For example any time correlation in the increments immediately demands a non-Markov property, since the increments are correlated.

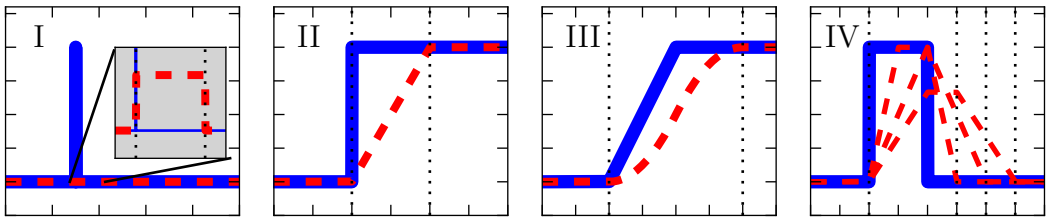


FIGURE 3.2: Effects of different kinds of irregularities on some measures averaged over a time window. The blue plots show some given quantitiy, the dashed red plots show the window-averaged value aligned to the right side of the time window. In I) one can see the effect of an outlier, leading to a rectangular form of the measure with the width of the time window. In II) one can see the effect of a sudden change of parameter, leading to a gradual change of the measure with a width of the time window. In III) one can see the effect of a linear change of parameter, leading to some gradual change in measure. The width of this transition is given by the sum of the width of the time window and the width of the transition. In IV) one can see the effect of a temporary change of parameter. The measure has some triangular form, depending on the size of the time window compared to the size of the temporary change.

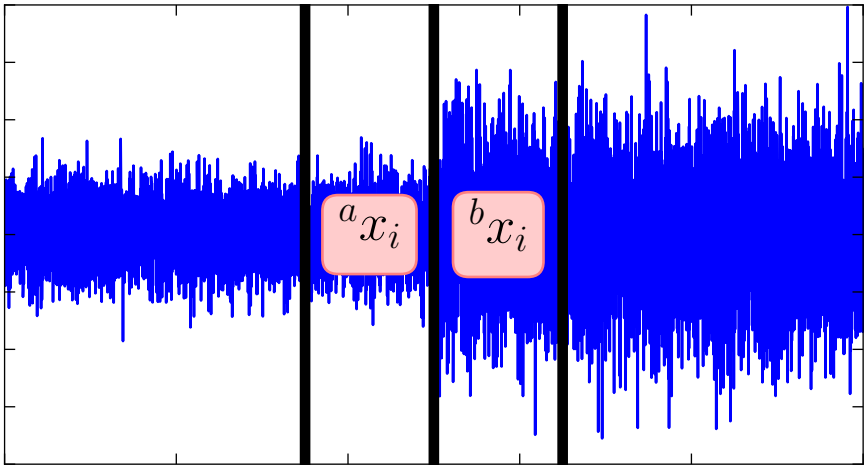


FIGURE 3.3: A time series with two sections of different stochastic properties. The transition is immediate. Measures can be taken of the $^{a}N$ datapoints of $^{a}x_i$ and the $^{b}N$ datapoints of $^{b}x_i$.

## 3.1 Central statistical moments

The central momentens of chapter 2.2.3 can be calculated for a data set as

$$\mu_2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{3.5}$$

for the variance,

$$\mu_3 = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{x_i - \mu}{\mu_2} \right)^3 \tag{3.6}$$

for the standardised empirical skewness and

$$\mu_4 = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{x_i - \mu}{\mu_2} \right)^4 \tag{3.7}$$

for the standardised empirical kurtosis.

## 3.2 Autocorrelation & AR(n)-processes

The autocorrelation as in chapter 2.2.4, can be used to construct different kinds of measures. The simplest one is the $a_1$-coefficient [1], that is the lag-1 value of the autocorrelation:

$$x_{k+1} = a_1 \, x_k + \epsilon_k. \tag{3.8}$$

If one regards the OU process (2.3.8) one can associate (in this linear case) (eq 2.46) a big $a_1$ with a strong restoring force and vice versa. The values that $a_1$ can take are on the interval $[-1, 1]$. It is shown in [21] that the autocorrelation for any Markov process follows an exponential decay law. If a process follows an exponential decay law, this can be used as measure to compare the strength of the exponential decay in different time windows.
It can also be used as a non-sufficient test for non-Markov property. For example one can measure the slope at lag-0 or find the position of an inflection point in the case of non-exponential decay.

## 3.3 Jensen Shannon & Kullback Leibler divergence

A way to search for a change in stochastic properties in data, is to compare the pdfs from different time windows. For this one needs a measure to determine the distance between two pdfs and one needs a reference pdf to compare the current pdf to. Such a measure can be defined with the help of the information entropy (chapter 2.2.5). The Kullback-Leibler divergence of two pdfs $p$ and $q$ is defined as

$$D_{\mathrm{KL}}\left(P||Q\right) := \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \, \mathrm{d}x. \tag{3.9}$$

If one writes the integrant as

$$p(x)\log p(x) - p(x)\log q(x) \tag{3.10}$$

one can interpret the Kullback-Leibler divergence as the entropy of $p(x)$ minus the entropy of $q(x)$ but weighted with the pdf of $p(x)$. A disadvantage of this measure is, that it is not positive definite and not symmetric

$$D_{\mathrm{KL}}(P||Q) \neq D_{\mathrm{KL}}(Q||P). \tag{3.11}$$

To eliminate these problems one can use the Jensen-Shannon divergence

$$\mathrm{JSD}(P||Q) = D_{\mathrm{JS}}(P||Q) = \frac{1}{2} D_{\mathrm{KL}}(P||M) + \frac{1}{2} D_{\mathrm{KL}}(Q||M) \tag{3.12}$$

with $M = \frac{1}{2}(P+Q)$. For both divergences $D(P||P) = 0$ applies; that means two identical distributions have a distance of zero. If one wants to apply the JSD to data one has to first calculate the pdfs (for example as in chapter 2.4.2 with an Epanechnikov kernel). As $\lim_{p\to 0} p\log p = 0$ there will not be any problems calculating this quantity, but since $\lim_{q\to 0} p\log q \to -\infty$ given $p \neq 0$ one has to remove data with very small $q$ and significantly bigger $p$.
There will be two different versions of the JSD used in chapter 4. The first one uses the pdf of the first time window as reference pdf. This way one compares the non-standardised stochastic qualities of the different time windows. The second version uses standardised pdfs (rescaled with eq 2.7), with the reference pdf being N(0,1). This evaluates how much the pdf differs from being a normal distribution.

## 3.4 Drift and diffusion estimation

In chapter 2.3.7 the Kramers Moyal expansion was presented. Equation 2.40 can be used to calculate drift and diffusion of a time series. Some stochastic process $X_t$ with time increment $\Delta t$ has the following drift ($n = 1$) and diffusion ($n = 2$) estimation at $x_k$:

$$\tilde{D}^{(n)}_{\Delta t}(x_k) = \frac{(x_{k+1} - x_k)^n}{n!\,\Delta t}, \tag{3.13}$$

with

$$D^{(n)}_{\Delta t}(x_k) = E\left[\tilde{D}^{(n)}_{\Delta t}(x_k)\right]. \tag{3.14}$$

Note that the averaging from equation 2.40 has been dropped and $\tilde{D}^{(n)}_{\Delta t}(x_k)$ is stochastically distributed. To further use this data one can utilise two different approaches:

- 1) Partitioning the value range and calculating the drift/diffusion in the respective intervals,

- 2) Calculating measures directly over all pairs of variates $(x_k|\tilde{D}^{(n)}_{\Delta t}(x_k))$, indirectly averaging them.

While way 1 directly reveals the functional correlation of drift/diffusion, way 2 is stochastically distributed. Thus, to obtain an idea of the actual functional relation way 1 should be preferred. To apply measures way 2 yields a better statistics, but weighs points near the equilibrium more than points far from it, thus it does not treat characteristics of phase space equally.

### 3.4.1 Finite time effects

Until now we assumed sufficiently small time steps in the calculation of the Kramers Moyal expansion. In experimental data the size of the time steps in conjunction with the magnitude of noise and drift can lead to significantly different results. To illustrate this, one can take a look at the drift and diffusion for Ornstein Uhlenbeck process with

$$D^{(1)}(x) = -\gamma x, \tag{3.15}$$

$$D^{(2)}(x) = D. \tag{3.16}$$

The finite time drift and diffusion of this Ornstein Uhlenbeck process was taken from [27]. The subscript $\tau$ is indicating a finite time sampling of the process (3.15), (3.16).

$$D^{(1)}_\tau(x) = -\frac{x}{\tau}\left(1 - e^{-\gamma\tau}\right) := -\gamma_\tau x \tag{3.17}$$

$$D^{(2)}_\tau(x) = \frac{1}{2\tau}\left[x^2\left(1 - e^{-\gamma\tau}\right)^2 + \frac{D}{\gamma}\left(1 - e^{-2\gamma\tau}\right)\right]. \tag{3.18}$$

One can see in (eq .3.17), that finite time effects lead to a $D^{(1)}_\tau$, that is also linear in $x$, with $\gamma_\tau = \frac{1-e^{-\gamma\tau}}{\tau}$ being the measured slope. As

$$\gamma = \frac{-\ln\left(1 - \tau\gamma_\tau\right)}{\tau}, \tag{3.19}$$

one must require

$$\gamma_\tau \leq \frac{1}{\tau}. \tag{3.20}$$

This means that there is an upper limit for the measured slope. The limit $\gamma_\tau \to \frac{1}{\tau}$ applied to (eq. 3.19) yields $\gamma \to \infty$, what corresponds to a white noise process. In the case of given $\gamma_\tau$

$$D^{(2)}_\tau = \frac{\tau\gamma_\tau^2}{2}x^2 + C. \tag{3.21}$$

The only unknown quantity here is $C$ and can be easily estimated with a least squares method (with $y_i$ being $\frac{(x_{i+1}-x_i)^2}{2\tau}$)

$$C = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \frac{\tau\gamma_\tau^2}{2}x_i^2\right). \tag{3.22}$$

Using eq 3.19 one can estimate:

$$D = -C\frac{\ln(1-\tau\gamma_\tau)}{2\tau^2\gamma_\tau - \tau^3\gamma_\tau^2}. \tag{3.23}$$

As last point we want to calculate $D_\tau^{(1)}(x)$ and $D_\tau^{(2)}(x)$ for a white noise process (e.g. uncorrelated Gaussian random numbers). Let

$$X_1 \sim \mathcal{N}(0, \sigma^2) \tag{3.24}$$

some Gaussian random variable with mean zero and variance $\sigma^2$, then one can simply (using Steiner translation theorem) write

$$D_\tau^{(1)}(x) = \left\langle \frac{x - X_1}{\tau} \right\rangle = -\frac{x}{\tau}, \tag{3.25}$$

$$D_\tau^{(2)}(x) = \left\langle \frac{(x-X_1)^2}{2\tau} \right\rangle = \frac{\sigma^2 + x^2}{2\tau} = \frac{1}{2\tau}x^2 + \frac{\sigma}{2\tau}. \tag{3.26}$$

This is consistent with above and $\gamma_\tau \to \frac{1}{\tau}$.

### 3.4.2 $D^{(1)}$-measures

The following methods to measure $D^{(1)}$ will be used:

- (lin): linear least squares fit over $\left(x_i|\frac{x_{i+1}-x_i}{\Delta t}\right)$
- (pf): pitchfork fit as used in 4.6 and 4.7
- (grad): the mean gradient of the Kramers-Moyal $D^{(1)}$ estimation: $\frac{1}{x_2-x_1}\int_{x_1}^{x_2}\mathrm{d}x\left|\partial_x D^{(1)}(x)\right|$ [28]
- (mean): the mean value of the Kramers-Moyal $D^{(1)}$ estimation

### 3.4.3 $D^{(2)}$-measures

The following methods to measure $D^{(2)}$ will be used:

- (mean1): the mean value over $\frac{(x_{i+1}-x_i)^2}{2\Delta t}$
- (offs): the offset as calculated in 3.22
- (min): the minimum of the Kramers-Moyal $D^{(2)}$ estimation
- (mean2): the mean value of the Kramers-Moyal $D^{(2)}$ estimation

### 3.4.4 Error estimation

An error estimation was used as in [27]. There, with the help of the conditional moments

$$M_{\tau}^{(n)}(x,t) = n!\tau D_{\tau}^{(n)}(x,t) \tag{3.27}$$

one can define with the help of a kernel estimator

$$\sigma_j^{(1)}(x_j) = \sqrt{\frac{M_{\tau}^{(2)}(x_j) - [M_{\tau}^{(1)}(x_j)]^2}{\sum_{k=1}^{N} \frac{1}{\Delta t} K(\frac{x_j - x_k}{\Delta t})}}, \tag{3.28}$$

$$\sigma_j^{(2)}(x_j) = \sqrt{\frac{M_{\tau}^{(4)}(x_j) - [M_{\tau}^{(2)}(x_j)]^2}{\sum_{k=1}^{N} \frac{1}{\Delta t} K(\frac{x_j - x_k}{\Delta t})}}, \tag{3.29}$$

as the standard deviations of drift and diffusion.

## 3.5 Markov time scale estimation

As mentioned in chapter 2.3.4 systems can have a certain 'memory time'. On a time scale smaller than this, the Markov property will not hold. To check if a process has such a time scale one can apply different methods.

This can be measures based on the autocorrelation, since the *acf* of a Markov process follows an exponential decay law. This can be used to check if a process is Markovian at all, or how much the *acf* differs from being exponential in decay.
Another method can be based upon the increment time series. If the increments of a process are correlated, the Markov property does not hold on the time scale of these increments. If a drift and diffusion estimation of the increment time series does not yield the white noise limit, one can assert, that the process is non-Markovian.
Another method employs the Chapman-Kolmogorov equation. For this, one compares both sides of equation 2.26 and measures the distance of these conditional probability distributions. This is done for all available time increments. The time increment for which the Chapman-Kolmogorov equation is fulfilled best, is chosen as Markov time scale. Thus one evaluates

$$S = \left\| p\left(x_1, t_1 | \, x_3, t_1 - 2\Delta t\right) - \int p\left(x_1, t_1 | \, x_2, t_1 - \Delta t\right) \, p\left(x_2, t_1 - \Delta t | \, x_3, t_1 - 2\Delta t\right) \, \mathrm{d}x_2 \right\| \tag{3.30}$$

for arbitrary $t_1$, where $\|\cdot\|$ stands for the chosen distance measure. This measure can for example be the sum of squares or the Jensen-Shannon divergence.
The conditional distributions can be obtained with the help of a two-dimensional Nadaraya Watson estimator by estimating $p\left(x_a, t_a; \, x_b, t_b\right)$ from the data set and

then using

$$p\left(x_a, t_a \middle| x_b, t_b\right) = \frac{p\left(x_a, t_a;\, x_b, t_b\right)}{\int p\left(x_a, t_a;\, x_b, t_b\right) \mathrm{d}x_a}. \tag{3.31}$$

The Markov time scale is that value of $\Delta t$, 'for which $S$ vanishes or is nearly zero (achives a minimum)'[4]. Their analysis of seismic data showed that $t_M$ increases before an earthquake.

# Chapter 4

# Applications

In this chapter the methods of the previous chapter will be applied to different data sets. Unless stated otherwise, all measures will be applied in sliding time windows as explained in Chapter 3. These data sets are firstly some synthetic time series to test the methods

- a pitchfork bifurcation with constant additive noise,
- the Haken-Zwanzig system with constant additive noise.

Secondly there are two time series of the 1996 Western North America blackouts. These show the bus voltage frequencies over some time up until the point of outage. These are

- a time series examined in [1], obtained from the author Christopher Danforth, called CD data set,
- a time series obtained from the 'Bonneville Power Administration'[1], called BV data set.

## 4.1 Supercritical pitchfork bifurcation with constant additive noise

The stochastic differential equation in this case reads as

$$\mathrm{d}x(t) = \left(\epsilon x(t) - x(t)^3\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{4.1}$$

To apply the concept of 'critical slowing down' as described in Ch2.X, the analytical solution to the deterministic part of (4.1) is obtained:

$$x(t) = \pm\frac{\sqrt{\epsilon}e^{\epsilon(t+C)}}{\sqrt{e^{2\epsilon(t+C)}-1}};\ C, \epsilon \in \mathbb{R}. \tag{4.2}$$

As $\epsilon \to 0$, the decay of $x(t)$ becomes a 'much slower algebraic function of t' [24]

$$x(t)_{\epsilon\to 0} = \frac{1}{\sqrt{2(t+C)}}. \tag{4.3}$$

[1] https://www.bpa.gov

As 'the system relaxes to equilibrium much more slowly than usual' [24], it is expected that the variance will increase near the bifurcation point $\epsilon = 0$. Expressed in a lax way the $a_1$-parameter describes the average relative increase of $|x_i|$ on time increments of lag-1. Thus an increase of this quantity can also be expected.
To apply the Fokker-Planck equation we want to assume, that the system has yet to be scaled (as one would have to with real data),

$$\mathrm{d}x_{scaled}(t) = \left(\epsilon\,(ax - c) - (ax - c)^3\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{4.4}$$

Because we assume window-wise stationary processes, one can always rescale with $x_{rescaled} = x - \bar{x}$. Thus one can always set $c = 0$ if one shifts the time series by its mean. If one now sets $a^* = a^{1/3}$ and $\epsilon^* = \epsilon\,a^{1/3}$, one gets

$$\mathrm{d}x_{unscaled}(t) = \left(\epsilon^* x - a^* x^3\right)\mathrm{d}t + \sigma\,\mathrm{d}W_t, \tag{4.5}$$

and one can easily obtain $\epsilon^*$ and $a^*$ by simple linear least squares method with

$$a^* = \frac{\sum x_i^4 \sum x_i y_i - \sum x_i^2 \sum x_i^3 y_i}{\left(\sum x_i^4\right)^2 - \sum x_i^2 \sum x_i^6}, \tag{4.6}$$

$$\epsilon^* = \frac{\sum x_i^4 \sum x_i^3 y_i - \sum x_i^6 \sum x_i y_i}{\left(\sum x_i^4\right)^2 - \sum x_i^2 \sum x_i^6}. \tag{4.7}$$

Since we used constant additive noise, the deterministic part of (4.1) is equal to $D^{(1)}$ (eq. 2.37). Thus we can obtain $\epsilon$ by applying (4.6), (4.7) either to $D^{(1)}$ or to $(x_i, \frac{x_{i+1}-x_i}{\Delta t})$ directly. The first method will neglect the higher relative frequency of datapoints near $x = 0$.
Figure 4.1 shows the simulation and evaluation of a pitchfork bifurcation as in eq. 4.1. For the simulation $10^6$ datapoints were used. The change of the bifurcation parameter was modelled by a hyperbolic tangens ranging form $-10$ to $10$. The time stepsize was $dt = 0.0001$ but for the evaluation only every tenth datapoint was used. The strength of the noise was taken as $\sigma = 0.2$. The blue plots in the figures are the respective evaluations, while the red plots are averaged values over 10 simulations to obtain smoother results.
The measures variance, $a_1$, pitchfork fit ($\epsilon^*$ from eq. 4.7) and linear fit (of $D^{(1)}$) all show the same behaviour - emulating the control parameter up until the point of bifurcation. The measure pitchfork fit follows the value of $\epsilon$ of equation 4.1 exactly, as $a = 1$. The linear fit, being an approximation of the pitchfork fit near $x = 0$ also follows the value of $\epsilon$. This is understandable, since due to the slowing down a flat linear part dominates the dynamics.
The Jensen Shannon divergence stays constant until shortly before the bifurcation point, where it drops abruptly. The location of the drop offs of the single simulations vary. This is due to the different outliers, that occur as the time series is becoming more volatile, thus this measure seems not to be reliable in this case.

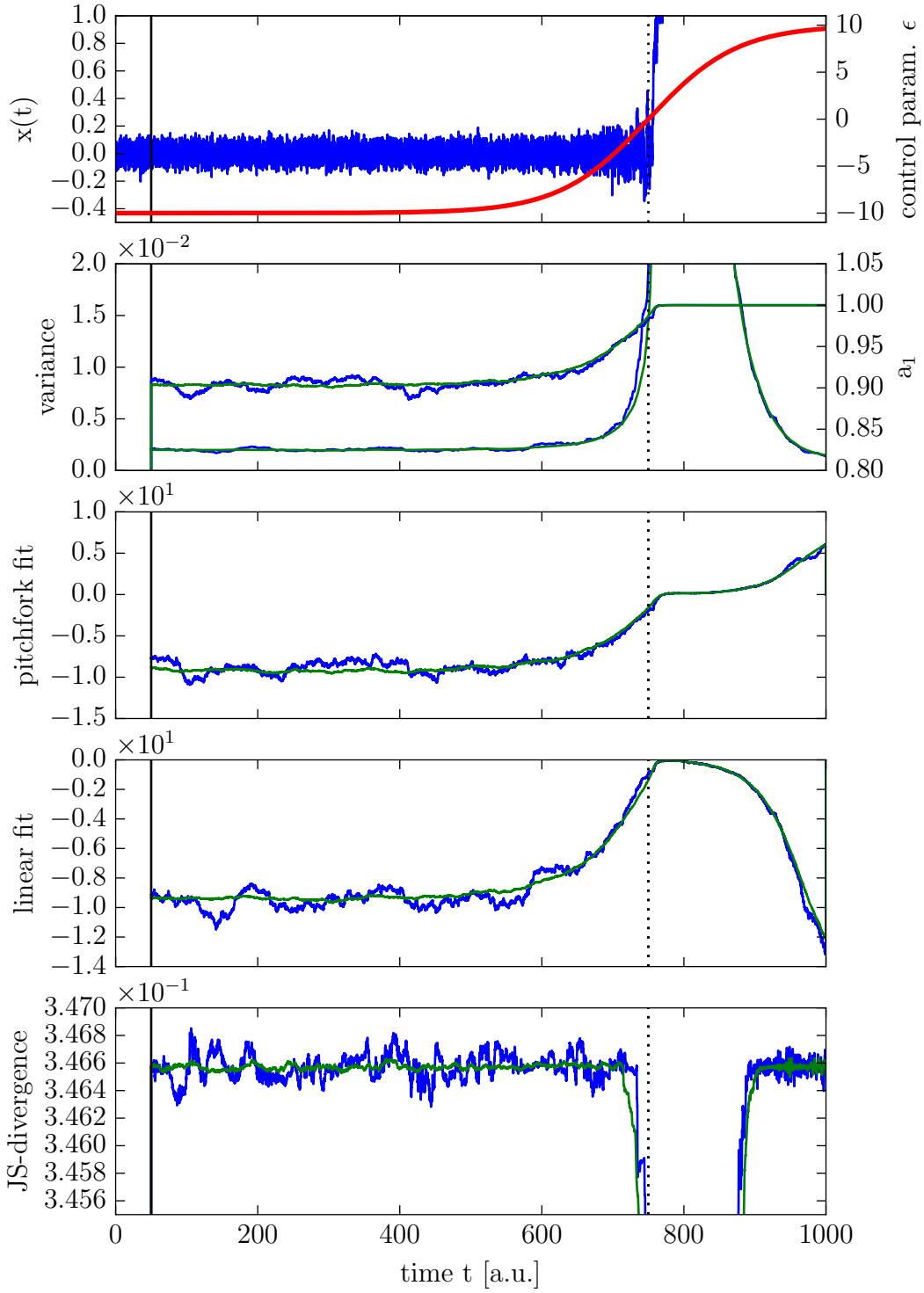


FIGURE 4.1: Supercritical pitchfork bifurcation. Dashed line shows the location of the bifurcation. In the second subplot, the upper plot is the $a_1$ estimation and the lower plot is the variance estimation. Blue plots are the estimations for one simulation. Red plots are the averages of ten simuations of the size of the blue ones.

## 4.2 Haken-Zwanzig system with constant additive noise

A very simple system with the qualities as described for equation 2.3, is the Haken-Zwanzig system. It has been used in [23] to demonstrate the concept of self-organisation with a system of only two dynamical variables. To examine the stochastical qualities of such a system, both variables were expanded with additive noise terms.

$$\begin{aligned} \mathrm{d}u(t) &= (\epsilon u(t) - u(t)s(t))\ \mathrm{d}t + \sigma_u\, \mathrm{d}W_t \\ \mathrm{d}s(t) &= (-s(t) + u(t)^2)\ \mathrm{d}t + \sigma_s\, \mathrm{d}W_t. \end{aligned} \tag{4.8}$$

For now we want to describe this system without noise. The variable $u$ stands for the dynamics of a force and the variable $s$ for the dynamics of a subsystem. In this situation the force is not independent of the subsystem, but reacts on the subsystem. To enforce the slaving principle we require $0 > \epsilon > -1$. Thus one can set $\dot{s} = 0$ and proceed as in equation 2.4. Thus $s$ is slaved by $u$. Using linear stability analysis for this system one can find fixed points for

$$\begin{pmatrix} u \\ s \end{pmatrix}_{stat,I} = \begin{pmatrix} 0 \\ 0 \end{pmatrix};\ \begin{pmatrix} u \\ s \end{pmatrix}_{stat,II} = \begin{pmatrix} \pm\sqrt{\epsilon} \\ \epsilon \end{pmatrix}. \tag{4.9}$$

For $\epsilon < 0$ (I) is stable and (II) is unstable. For $\epsilon > 0$ (I) is unstable and (II) is stable. The idea is now to examine different measures as the system with noise approaches its bifurcation point at $\epsilon = 0$ as it exhibits separation of time scales. For the slow variable $u$, variance and $a_1$ increase, indicating the signs of csd. The drift estimation via fit of a pitchfork bifurcation and via linear fit emulate the bifurcation parameter $\epsilon$. The Jensen-Shannon divergence exhibits no change as the bifurcation is approached. For the fast variable $s$, only the $a_1$-coefficients show an increase as the bifurcation gets approached.
For the simulation $10^7$ datapoints with timesteps of $\Delta t = 0.001$ were used, but only every tenth step was taken. For the calculations windows of the size of $50000$ steps were used.

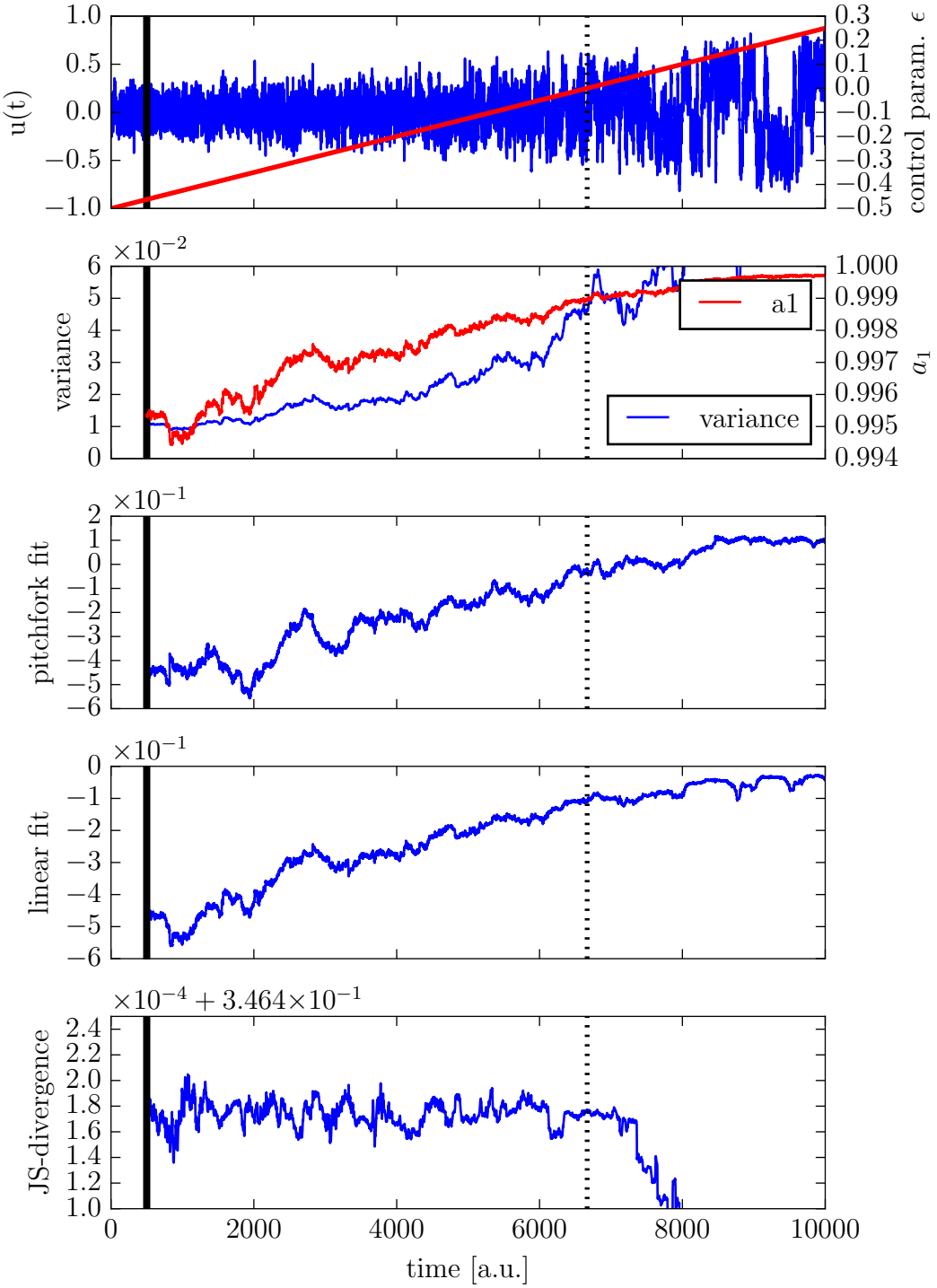


FIGURE 4.2: Slow variable $u$ of the Haken Zwanzig system. Dashed line indicates the point of bifurcation. First plot shows the simulated time series; the second plot the variance and $a - 1$-coefficients; the third plot shows a drift estimation by fitting of a pitchfork bifurcation; the fourth plot shows a drift estimation by linear fitting; the fifth plot shows the Jensen-Shannon estimation.

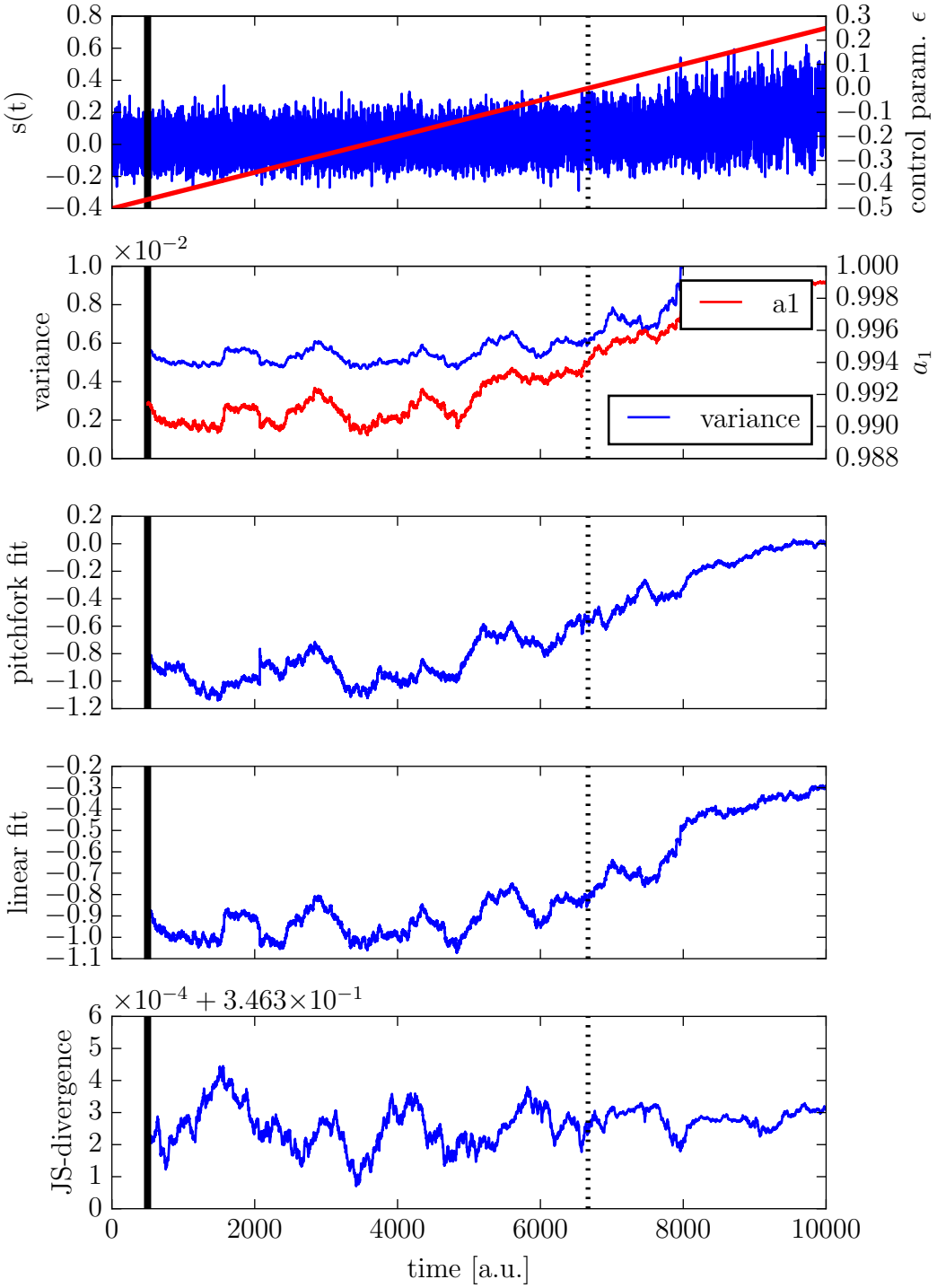


FIGURE 4.3: Fast variable $s$ of the Haken Zwanzig system. Dashed line indicates the point of bifurcation. First plot shows the simulated time series; the second plot the variance and $a-1$-coefficients; the third plot shows a drift estimation by fitting of a pitchfork bifurcation; the fourth plot shows a drift estimation by linear fitting; the fifth plot shows the Jensen-Shannon estimation.

## 4.3 Power outage data

In this chapter two time series of the '1996 Western North America blackouts' of August 10 1996 are used. One of these was obtained from Christopher M. Danforth as he analysed this time series in [1] (called CD from now on). The other time series was obtained from the Bonneville Power Administration (called BV from now on). Both time series stem from the same event, but no absolute time was provided; just time series 'up until the point of segregation'. To align the two time series temporally, one could on the one hand align them from the point of segregation, on the other hand one could try to find some distinctive features like peaks and align them originating from those points. In (fig. 4.4) one can see both datasets, aligned to the point of segregation. Those are the raw time series (upper subplot) and the GKS-detrended ($\sigma = 5\,s$) time series (lower subplot). In each plot the shorter time series belongs to the CD dataset.

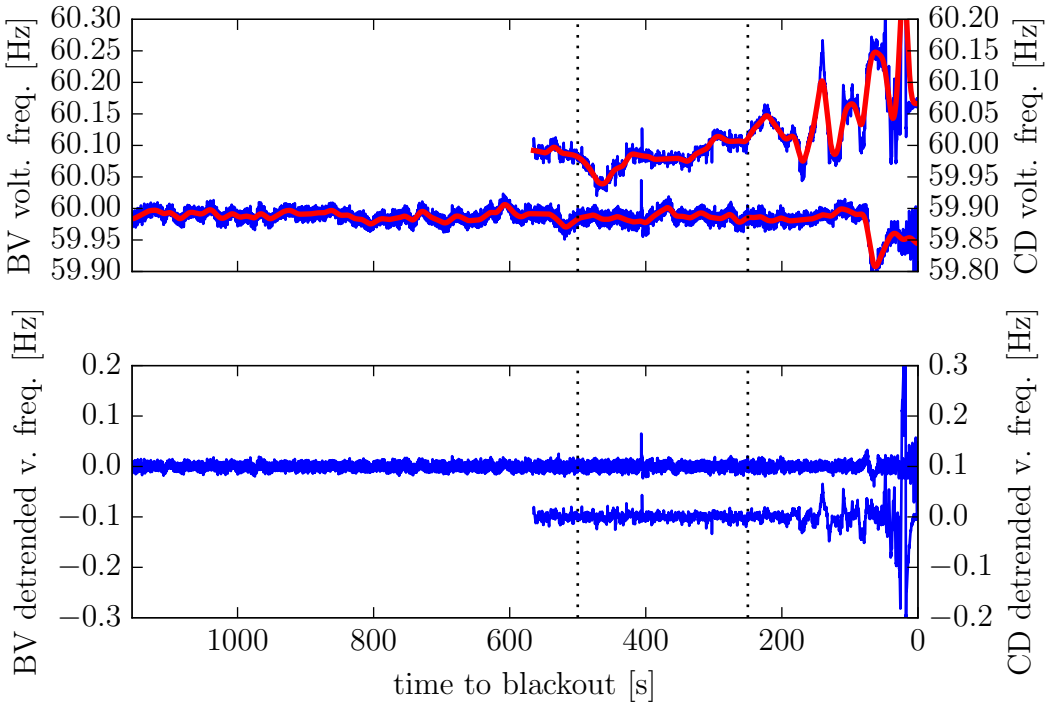


FIGURE 4.4: Both time series of the outage (blue) and their Gaussian kernel smoothed version (red) in the upper subplot. The detrended time series are depicted in the lower subplot. In each plot the shorter time series belong to the CD data set. Both are aligned to the point of segregation. Note the outliers in both time series at around $t = 400\,\mathrm{s}$. These are approximately $1\,\mathrm{s}$ apart.

Another possibility to align the two time series would be to calculate the covariance in dependence of some temporal shift, and to assign the shift of maximal covariance as the best match. The window of analysis is shown as dotted lines in fig. 4.4. It is important to note, that both time series have differently sized timesteps. Those are $\Delta t_{BV} = 0.05\,\mathrm{s}$ and $\Delta t_{CD} = 0.02415\,\mathrm{s}$. To make those

two time series compatible, the CD time series was mapped to the same time scale as the BV time series. This was done by taking the value with the smallest temporal difference of the CD time series compared to the BV time series. The temporal shift was defined as $t_{shift} = 0\,\mathrm{s}$ for an alignment at the point of segregation. A positive $t_{shift}$ corresponds to a shift of the BV time series positive in time. The time shift reaches from $t_{shift} = -10\,\mathrm{s}$ to $t_{shift} = +10\,\mathrm{s}$. The periodicity

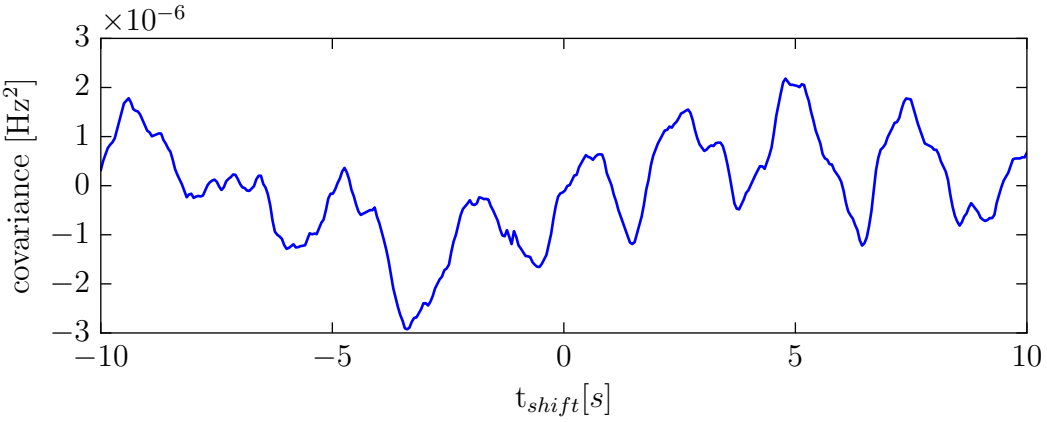


FIGURE 4.5: Covariance of the BV and CD datasets, applied in the time window depicted in fig 4.4 with black dotted lines, shifted by $t_{shift}$

in the covariance corresponds to a periodicity also found in the BV data, but no distincitve peak can be identified, thus no optimal shift can be obtained this way. Thus, the time series will always be aligned by the point of segregation from now on. Since we don't use any measures that compare the time series directly, an exact alignment (to the precision of a few time steps) is not needed. Moreover the possible error of $\sim 1\,s$ is very small in comparison with the used time windows of $\sim 250\,s$.

As next point we want to examine the autocorrelation functions of the data sets. In (fig. 4.6) one can see the autocorrelation function (acf) of the BV data set (blue dots). In the upper subplot one can see the acf of the first half of the GKS-filtered time series. The lower subplot shows the Fourier transform of the data. One can clearly see a peak at around $\sim 0.4\,[\mathrm{Hz}^{-1}]$, that corresponds to the periodicity seen in the upper autocorrelation (and the periodicity in the covariance). This part was damped to a level to match the stochastic properties of the surrounding data of the fourier transform. The resulting acf can be seen in the central subplot. This acf still shows periodic features in values of small time shift, while it is unclear if this occurs due to the detrending. To illustrate this, figure 4.8 shows the autocorrelation function of white noise, with the lower third of the frequencies of the signal set to zero. One can clearly see a negative autocorrelation value for short time shifts. Thus, removing low frequencies of a stochastic process, can lead to significant altered values of the acf for short time shift.

In (fig. 4.7) one can see the autocorrelation of the CD dataset in the upper plot

and its first (blue) and second (red) derivative in the central subplots. The Fourier transform in the lower subplot does not exhibit any frequencies that dominate the spectrum, thus no filter was applied. One can see, that a simple exponential fit does not deliver a satisfactory result. The first derivative is non-zero at $t_{shift} = 0$ and the second derivative shows the existence of an inflection point at around $t_{shift} = 0.2\,s$. To measure the change of the acf over time the gradient at zero time shift and the position of the inflection point were chosen. A shift of the inflection point towards $t_{shift} = 0\,s$ would indicate, that the Markov property could be fulfilled again, since the condition of an exponential acf could be fulfilled. This can be seen in figure 4.9.

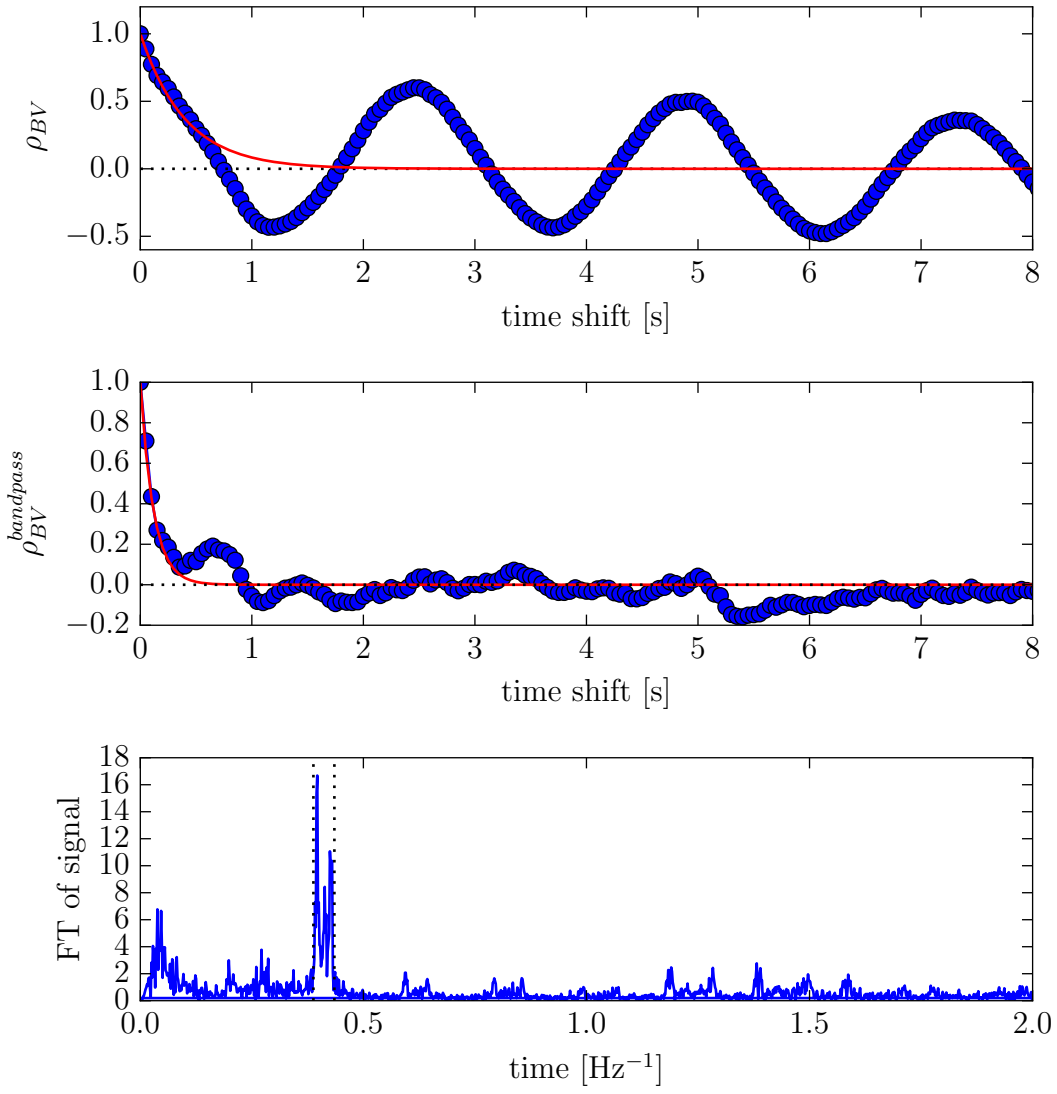


FIGURE 4.6: Autocorrelation of the BV data set. Upper subplot: acf of the unaltered time series; Central subplot: acf of the time series with periodicity removed in fourier space.; Lower subplot: Fourier transform of the time series. The dotted lines show the interval of dampened frequencies.

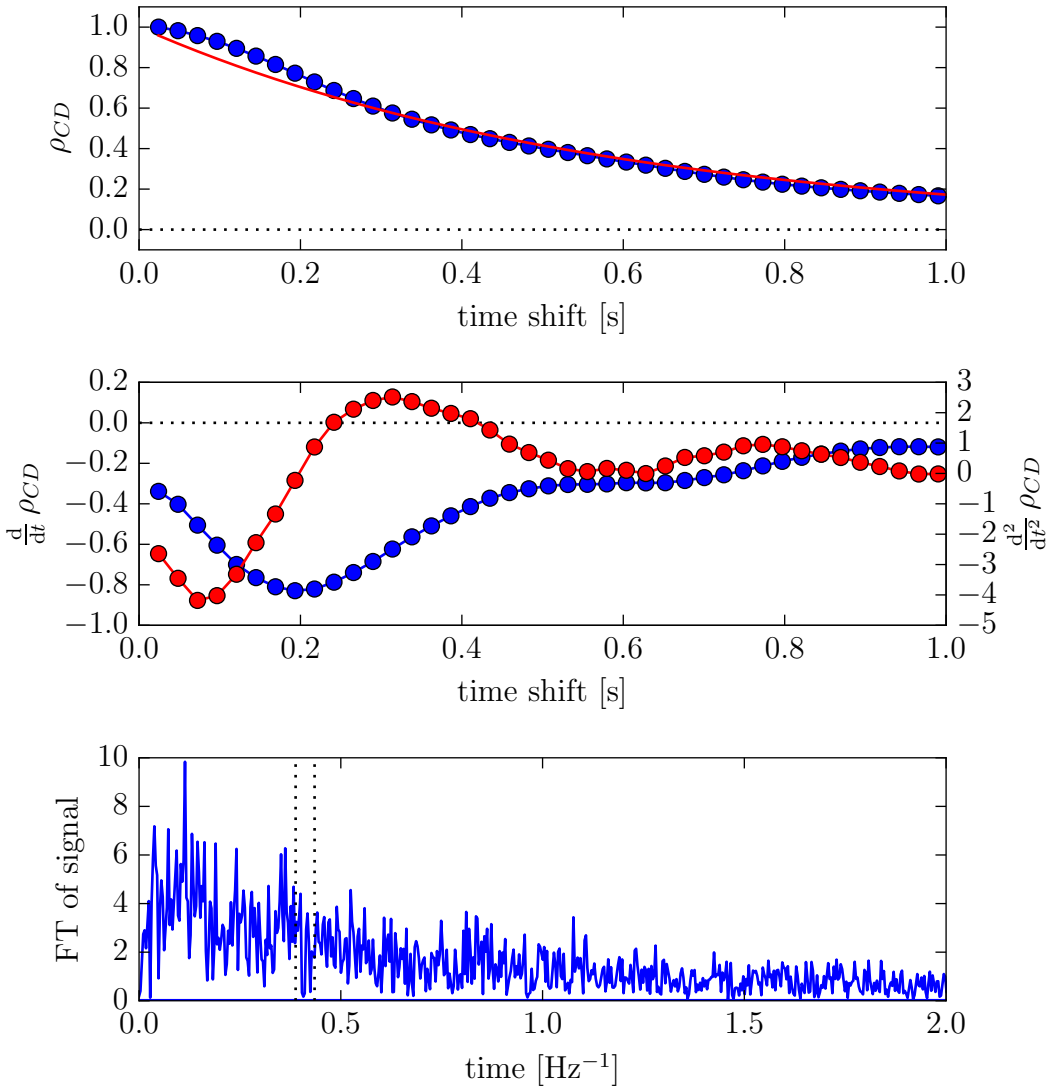


FIGURE 4.7: Autocorrelation of the CD data set. Upper subplot: acf of the time series; Central subplot: First(blue) and second(red) derivative of the acf; Lower subplot: Fourier transform of the time series.

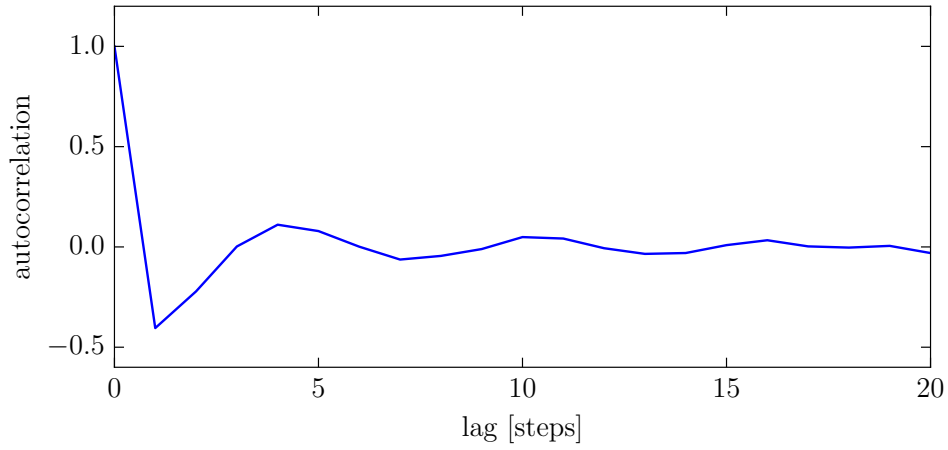


FIGURE 4.8: Autocorrelation function of white noise with the lower third of its frequencies set to zero. The acf is negative for short time shifts.

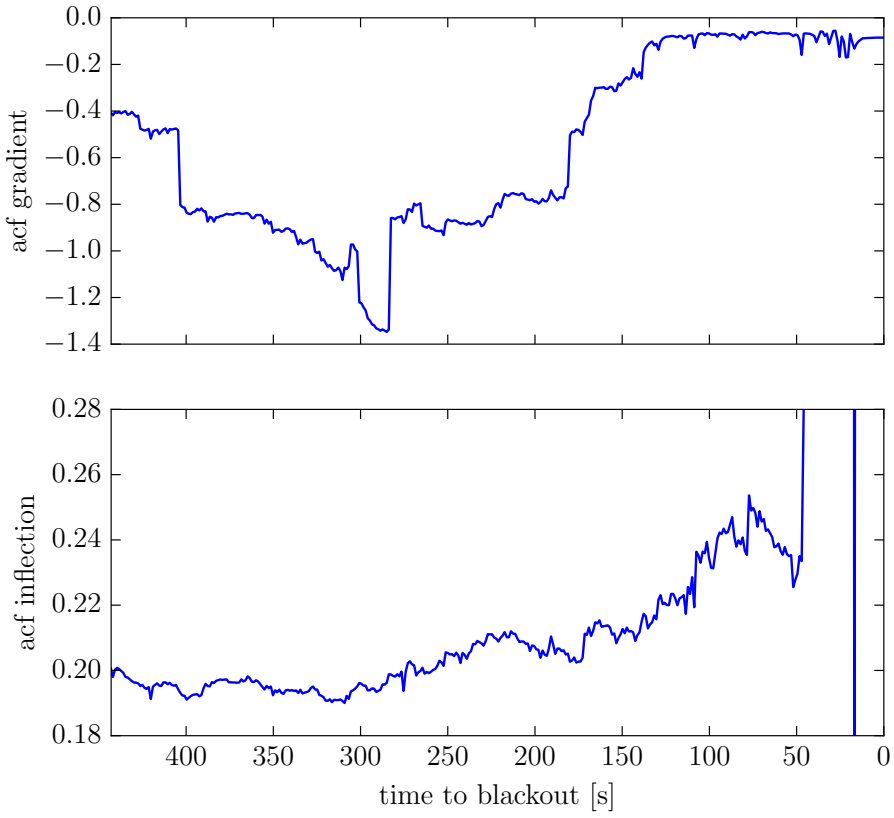


FIGURE 4.9: The gradient and inflection point of the acf of the CD data set. A window size of 5000 time steps was chosen. The value of the gradient rises at around $t = 175\,s$, being almost constant close to zero at $t = 140\,s$. The inflection point gradually shifts to the right over the time of the data set.

### 4.3.1 Statistic moments, $a_1$ and Jensen Shannon divergence

Up next we examine the stochastic moments, the $a_1$ coefficient and the Jensen-Shannon divergence for both time series and their increment time series. Differently sized time windows were used, but a size of 5000 steps yielded best results. The stochastic moments were applied as described in chapter 2.2.3. The $a_1$ coefficient was obtained using simple least squares method. The Jensen-Shannon divergence was calculated in the two different ways introduced in chapter 3.3. In the case of untransformed pdfs, the reference pdf chosen was that one of the first time window. In the case of the N(0,1) transformed pdfs, the reference pdf chosen was N(0,1). Black vertical lines in the big figures (4.10, 4.11, 4.14, 4.15) indicate the right side of the time windows used for the pdfs in the small figures (4.12, 4.13, 4.16, 4.17). There the plots for the unnormed pdfs - the black plots - are the pdf of that window, while the light blue plots show the pdfs of the other windows as comparison. The plots for the normed pdfs always show the $N(0,1)$ distribution as comparison as red plot.
Figure 4.10 shows the results for the BV dataset. The outlier at around $t = 400\,\mathrm{s}$ was removed. The time windows used to calculate the pdfs in (fig 4.12) are located at (I $= 905\,\mathrm{s}$, II $= 530\,\mathrm{s}$, III $= 390\,\mathrm{s}$, IV $= 80\,\mathrm{s}$). Measures closer to the point of segregation than $t = 80s$ will be neglected because of the volatility of even the smoothed time series from that point onward (a change can be directly seen in the time series).
The $a_1$ coefficient seems to be constant up until $t = 530s$. At that time point it linearly decreases up until $t = 80s$ with notches at $t = 405s$ and $t = 280s$. This suites case fig. 3.2.II, thus assuming jumps of the $a_1$ property at $t = 530s$ and $t = 405s$.
The variance exhibits jumps at $t = 400s$ and $t = 150s$. The temporal distance between these jumps fits the size of the time window. This suites case fig. 3.2.I, thus indicating a just very short increase in variance at around $t = 400s$.
The skewness ($\mu_3$) has no distinctive feature before $t = 80s$. The kurtosis ($\mu_4$) has a similar behaviour as the variance on the interval $(400, 100)$, but also linearly increases in this time frame. Thus suggesting some combination of 3.2.I&II; therefore indicating a jump in value initiated by a short peak.
The JSD has a gradual increase until $t = 400s$. At that point an increase of the width of the time window indicates case 3.2.I; thus showing a short peak of this measure at $t = 400s$. The corresponding pdfs are shown in figure 4.12. Pdf I, II and III are broader than pdf IV and show asymmetric features at the tip. Pdf III is more symmetric than pdfs I and II.
The normed JSD shows a drop off at $t = 400s$, followed by a gradual decrease of measure, indicating a sudden change in stochastic properties followed by a gradual change. In this case the pdf are nearing a normal distribution. As pdf I and II are bumpy, pdf III is smoother and pdf IV is almost Gaussian.

The time series of the CD data is characterized by more outliers, than the BV data (fig 4.11).
The $a_1$ measure reacts sensitive to these outliers and a behaviour as in fig. 3.2.I can be seen at $t = 400s$ to $t = 280s$ and $t = 310s$ to $t = 190s$. From around

$t = 140s$ onward, $a_1$ stays constant.
Around this point the variance increases up until the point of segregation over a range bigger than a time window, indicating a growth in value (as one can also see directly in the time series).
The skewness and kurtosis show a sudden increase in value at around $t = 150s$, with subsequent decline to a value around the starting value. Measures of statistical properties can be biased by too small time windows to evaluate the actual value. At this point the time series becomes more volatile thus making these measures unusable.
The normed JSD is subject to the same problem as the skewness/kurtosis (pdfs shown in figure 4.13).
The unnormed JSD has a behaviour very close to the variance. Pdf IV is flat in comparison with pdfs I, II and III, as one could have predicted by only looking at the time series.

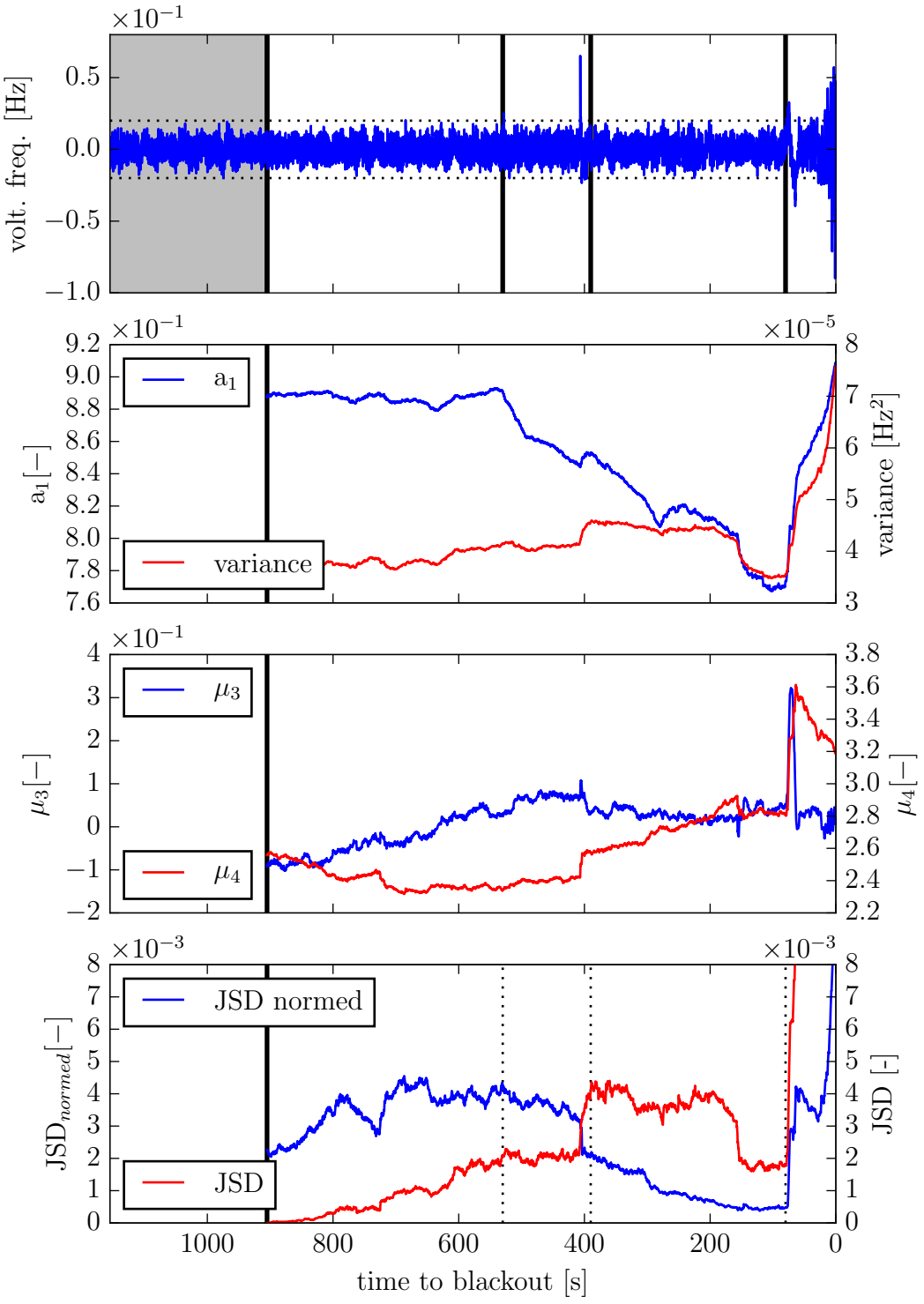


FIGURE 4.10: BV time series: statistic moments, $a_1$ measure and Jensen Shannon divergence

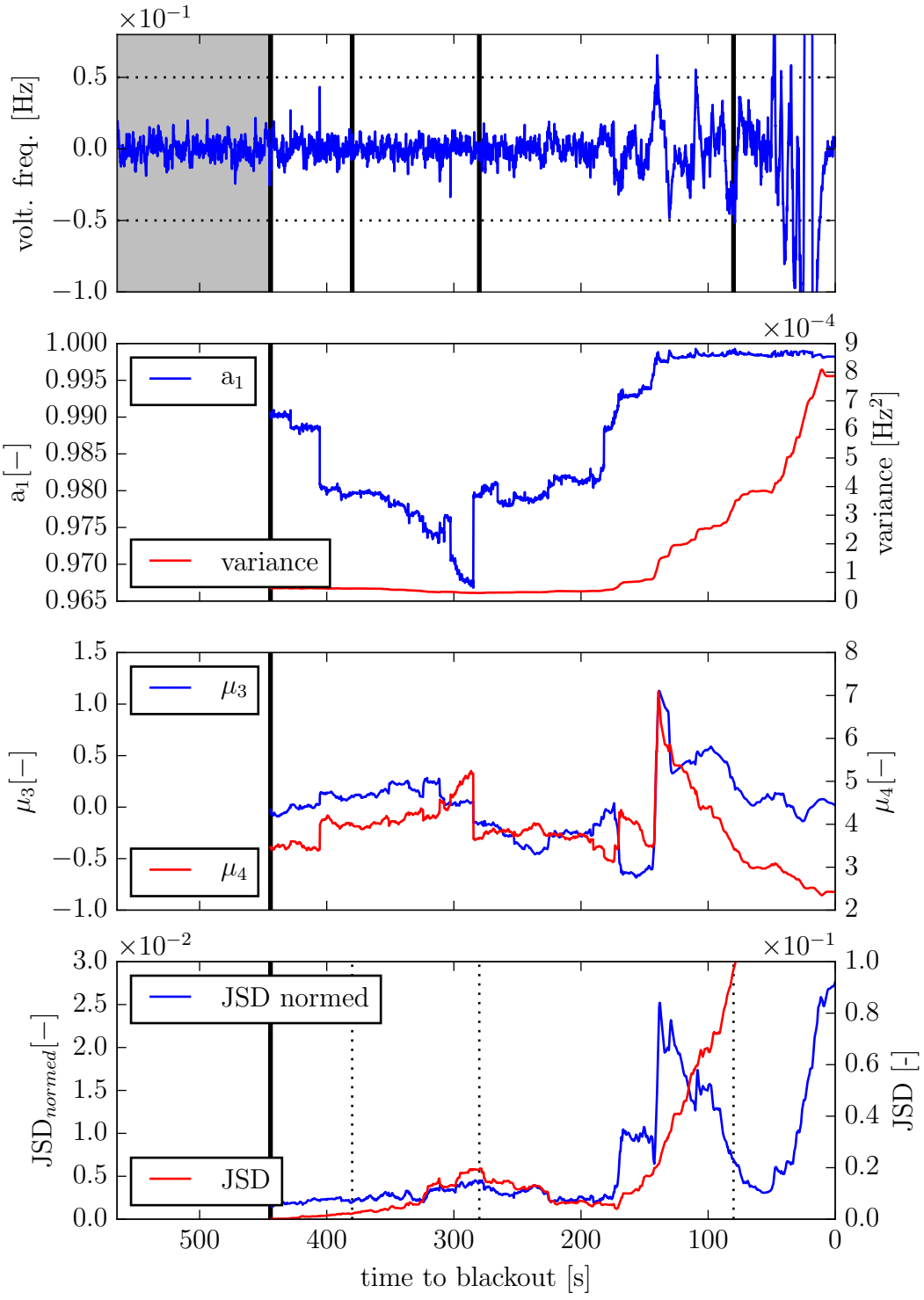


FIGURE 4.11: CD time series: statistic moments, $a_1$ measure and Jensen Shannon divergence

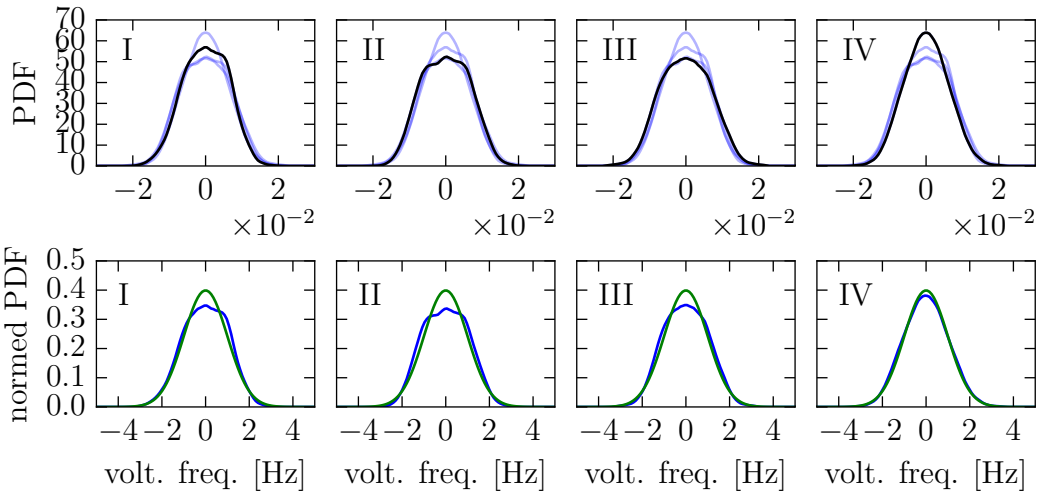


FIGURE 4.12: BV dataset: pdfs used for the JSD measures. The pdfs get smoother and more gauss-like as the time series advances.

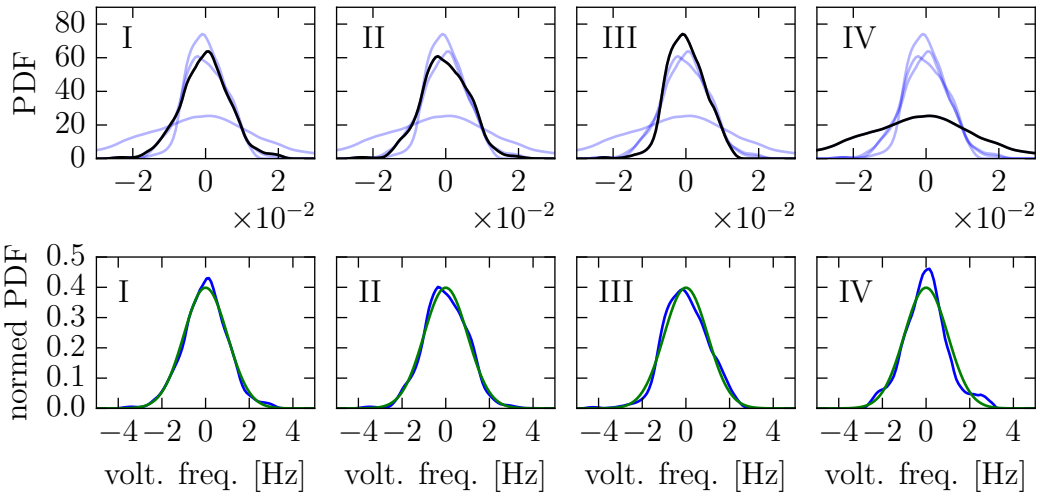


FIGURE 4.13: CD dataset: pdfs used for the JSD measures. The pdfs are much uneven, compared to the BV data set.

Figure 4.14 shows the evaluation of the increment time series of the BV data. The $a_1$ coefficient seems to be constant (slightly increasing from $-0.03$ to $0.03$ ) up until $t = 530s$. At this point the $a_1$ value drops to $-0.3$ up until shortly before the point of segregation. The gradual decrease over a period of $\Delta t = 510s$ indicates drops of this value over the time window beginning at $t = 530s$.
The variance is also almost constant until $t = 530s$, then evenly rising up until $t = 280s$. The duration of this increase corresponds to the time window, indicating a jump in variance at $t = 530s$.
The skewness and the kurtosis, as before, show jumps at $t = 530s$, again due to the standardisation of these measures.
The JSD again shows a behaviour very similar to that of the variance, indicating a sudden change in stochastic properties at $t = 530s$. The pdfs are shown in figure 4.16. Pdf IV is flat in comparision with pdfs I, II and III.
The normed JSD has an increased value over the interval $(530, 300)$. This is again due to this being a normed measure.

Figure 4.15 shows the evaluation of the increment time series of the CD data set. The $a_1$ coefficient drops right at the beginning at $t = 420s$ until $t = 300s$, at which point it stays within a certain range. The range of this decrease again is the width of the time window, indicating a jump in value at $t = 420s$.
The variance is almost constant up until $t = 100s$. At this point the variance increases, as one can also see directly in the increment time series.
The skewness and kurtosis have gradual growth/decrease over the interval of analysis, yielding no information.
The JSD and normed JSD both exhibit a behaviour close to that of the variance. The pdfs can be seen in figure 4.17.

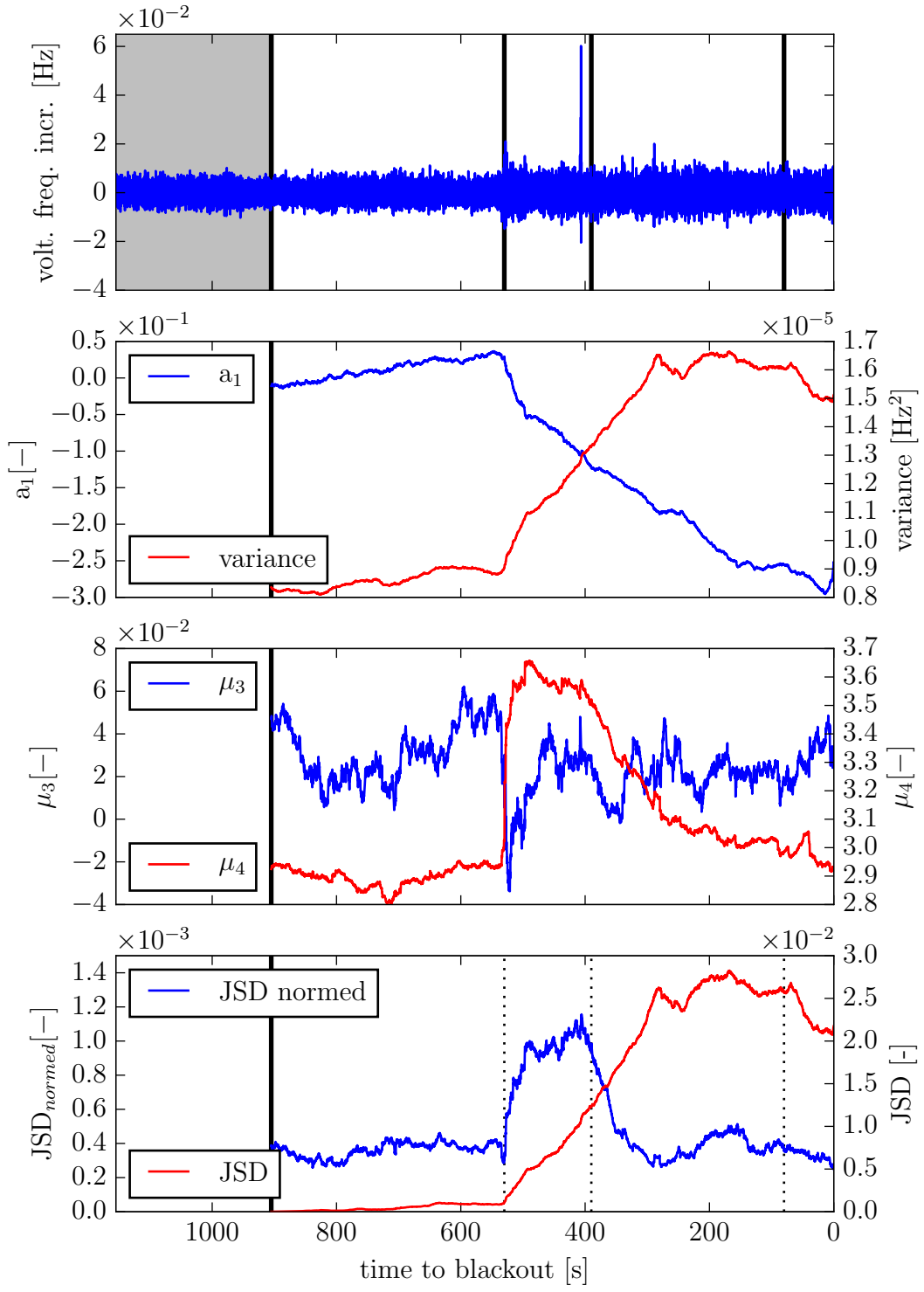


FIGURE 4.14: BV increment time series: statistic moments, $a_1$ measure and Jensen Shannon divergence

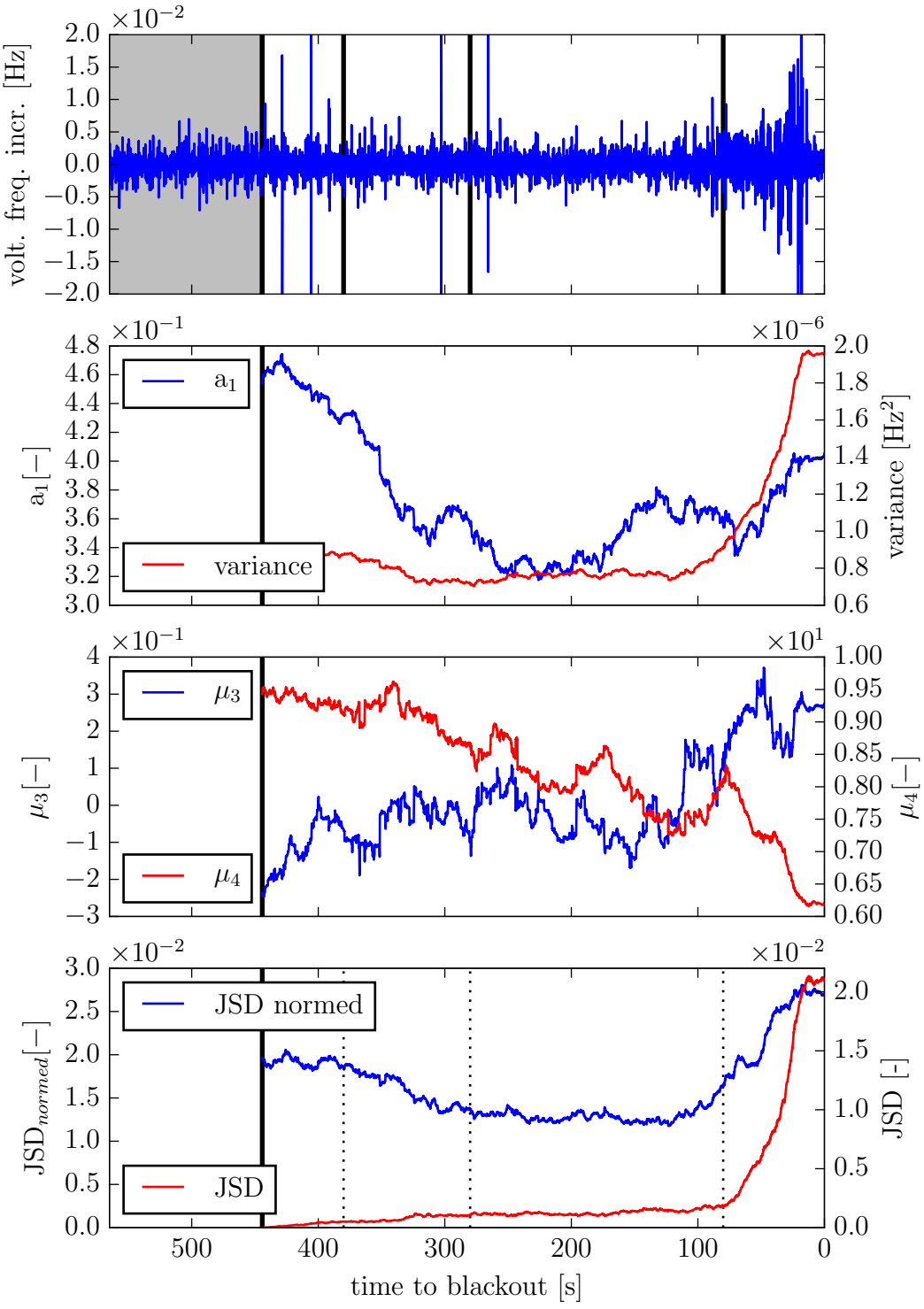


FIGURE 4.15: CD increment time series: statistic moments, $a_1$ measure and Jensen Shannon divergence

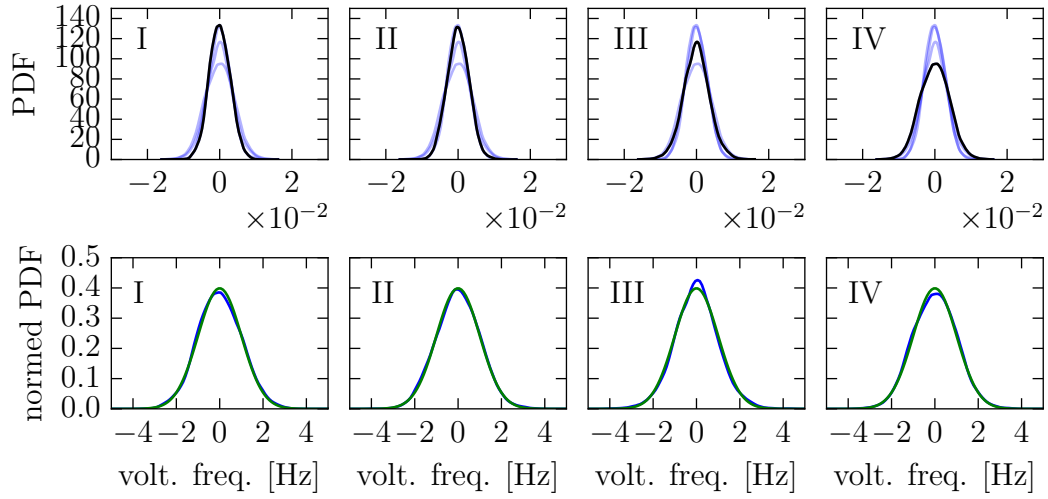


FIGURE 4.16: BV increment dataset: pdfs used for the JSD measures. While the unnormed pdfs get broader as the time series progresses, the normed pdfs always stay very close to the $N(0,1)$-distribution.

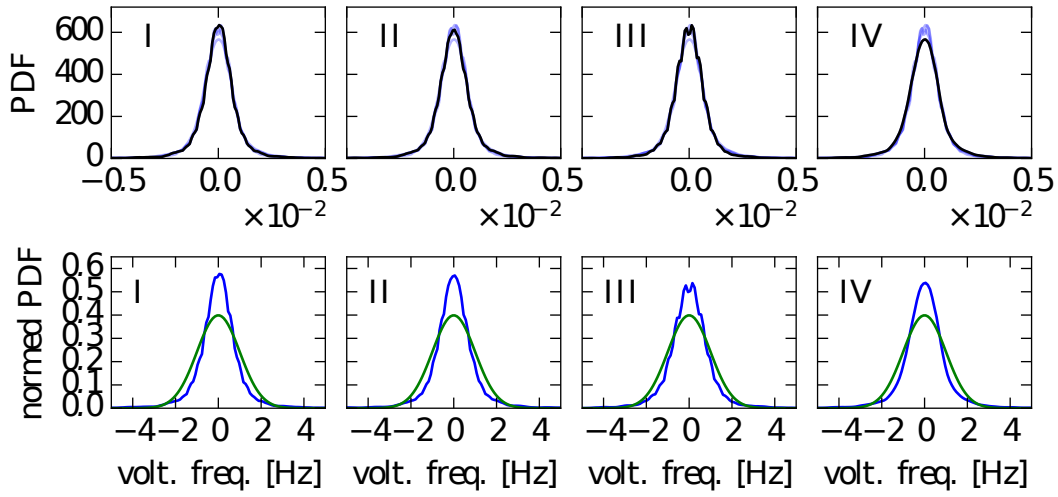


FIGURE 4.17: CD increment dataset: pdfs used for the JSD measures. While the pdfs seem to maintain their general form, the last pdf is smoother then the other ones. The normed pdfs are clearly non-gaussian.

### 4.3.2 Drift & diffusion estimation

Next we examine the (increment) time series with some measures based on drift & diffusion estimation. The scheme of the following figures (4.19, 4.20, 4.21, 4.22) repeats itself. In the first subplot the time series and its GKS-detrended version is depicted. In the second subplot different measures to estimate the drift ($D^{(1)}$) are shown. In the third subplot different measures to estimate the diffusion ($D^{(2)}$) are shown. These drift and diffusion measures are all scaled to the interval $[0, 1]$, to make them easily compareable. In the lower two rows the Kramers-Moyal estimation for drift and diffusion are shown window-wise. The position of these time windows in the time series is marked by the vertical lines in the upper plots. The names of different measures are explained in chapters 3.4.2 & 3.4.3. For the calculation of these measures some preprocessing was done. Outliers were removed, the measures where only estimated on an interval containing suffcient datapoints and a lag of two data points was used. Errors for the $D^{(1)}$ and $D^{(2)}$ were calculated as in 3.28 and 3.29, but were too small to display in these plots.

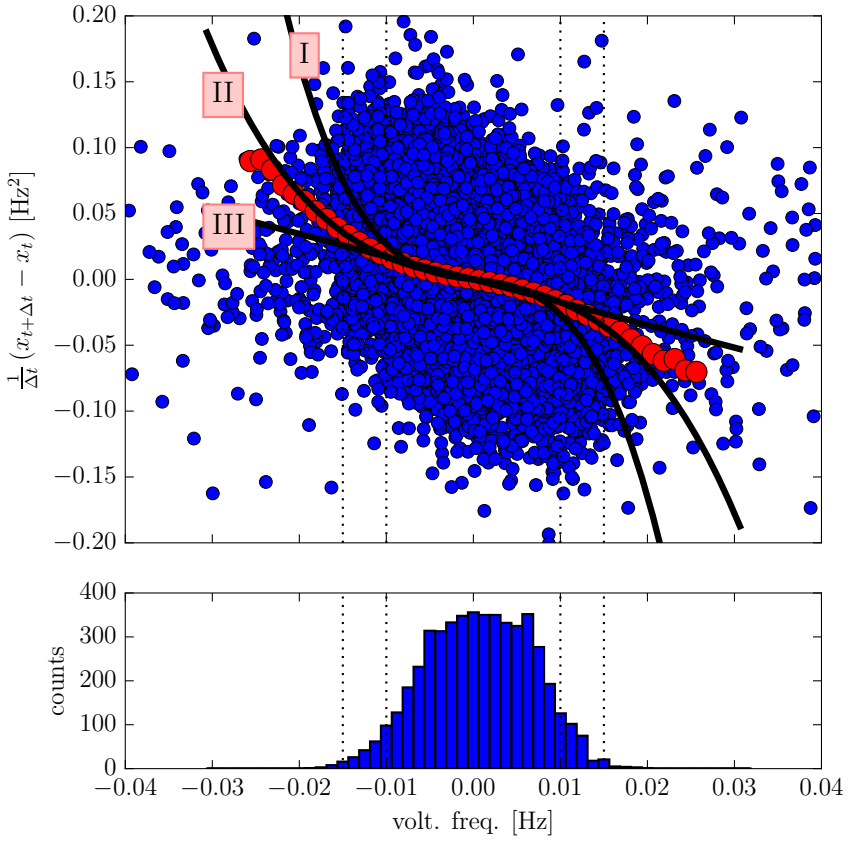


FIGURE 4.18: Examplary plot for the following drift estimations. Upper plot shows the point cloud used (blue) to calculate the $D^{(1)}$-coefficients (red). The black plots show the pitchfork fits on the interval $[-0.01, 0.01]Hz$(I) and $[-0.015, 0.015]Hz$(II), and a linear fit (III).

Figure 4.19 shows the evaluation of the BV time series. Measures were calculated on the datapoints in the interval $[-0.01, 0.01]\,\mathrm{Hz}$ (marked by dotted line). The red parabolas in the diffusion plots show the diffusion estimation as in 3.21. This was done to show the effect of finite time sampling (the actual diffusion is near constant).
All diffusion measures exhibit the same behaviour. There is almost no increase up until $t = 530\,\mathrm{s}$. From that point onward there is a gradual increase up to $t = 280\,\mathrm{s}$. This again is exactly the width of the time window, indicating a jump in diffusion at $t = 530\,\mathrm{s}$. There are also small features at $t = 400\,\mathrm{s}$ and $t = 150\,\mathrm{s}$ origninating from the peak at $t = 400\,\mathrm{s}$.
The drift measures all have similar behaviour but (pf) differs a bit. The measures (lin), (grad) and (mean) are almost constant until $t = 530s$. From that point they decrease until $t = 280s$, indicating a jump in value at $t = 530s$. The (pf) measure slightly increases up until $t = 400s$ and then decreases until $t = 150s$, indicating a jump at $t = 500s$.

Figure 4.20 shows the evaluation of the CD time series. Measures were calculated on the datapoints in the interval $(-0.X, 0.X)$. The red parabolas in the diffusion plots again show the diffusion estimation as in 3.21.
The diffusion measures suffer from the bad quality of this dataset. There are indents in all measures, except (mean2). This measure is almost constant up until $t = 100s$, at that point it increases rapidly up until the point of segregation.
The drift measures differ from each other at this data set. (pf) has no distinctive features. The features of (grad) cannot be associated with any behaviour and appear to be just random. The measures (lin) and (mean) behave similarly. At around $t = 190s$ to $t = 70s$ the measure increases, indicating a jump at $t = 190s$.

Figure 4.21 shows the evaluation of the increment time series of the BV data set. The cutoff was set to the interval $(-0.003, 0.003)\mathrm{Hz}$. The differrent drift and diffusion are again very similar to each other. The black parabolas in the diffusion plots, this time show the white noise limit for the diffusion estiamtion (eq 3.26)
The drift is almost constant up until $t = 530s$, at what point it drops over a period of $350\,\mathrm{s}$, indicating a drop over a time window of $130\,\mathrm{s}$. The slope of the linear fit were all near the white noise limit of the OU-process.
The diffusion is almost constant until $t = 530s$ at which point it increases until $t = 280s$, indicating a jump in diffusion at $t = 530s$ in all measures.

Figure 4.22 shows the evaluation of the increment time series of the CD data set. The cutoff was set to the interval $(-0.0015, 0.0015)\mathrm{Hz}$. The red parabolas in the diffusion plots again show the diffusion estimation as in 3.21, indicating that a near-constant diffusion.
The drift measure (grad) again has a very low quality. The measures (pf) and

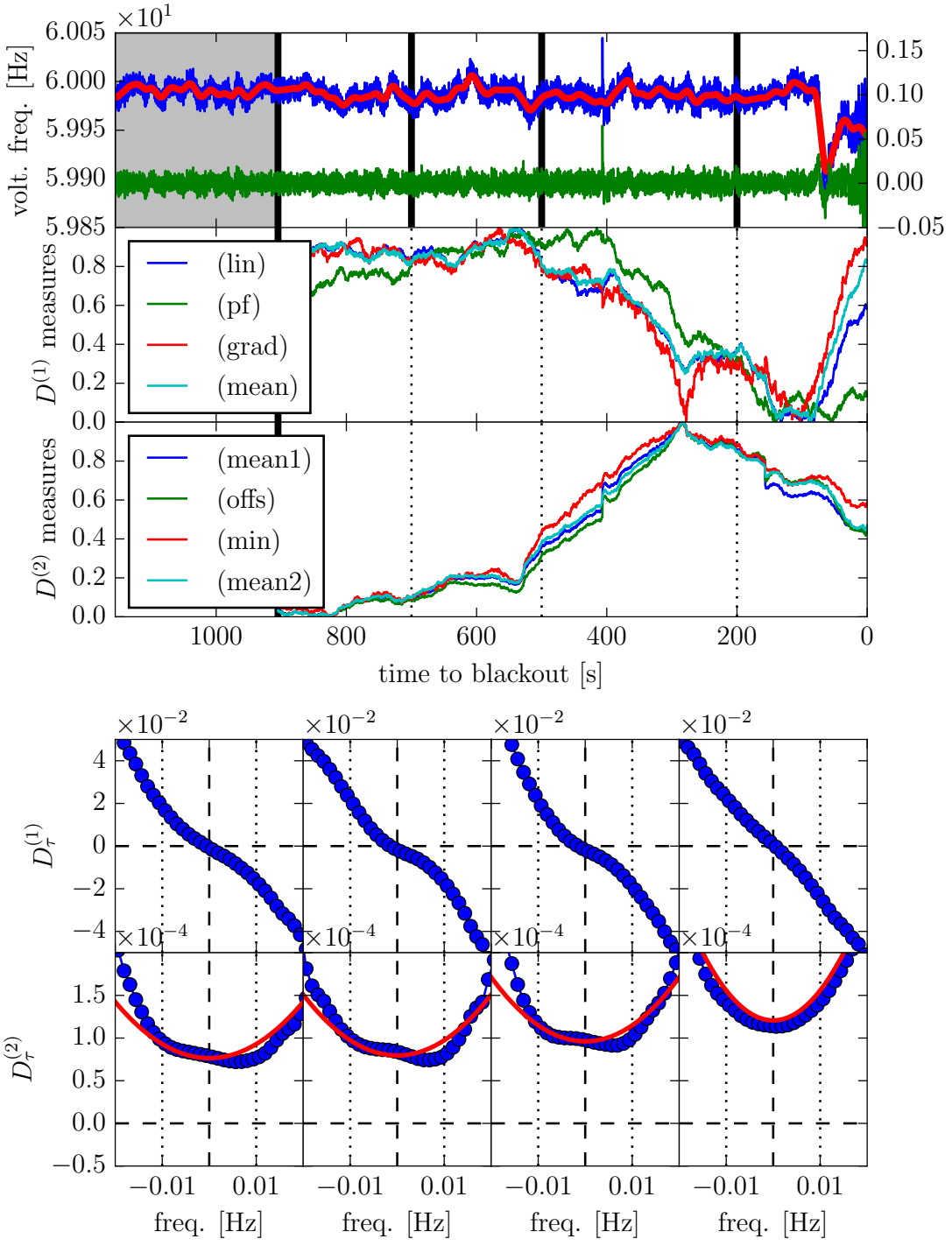


FIGURE 4.19: Drift & Diffusion estimation for the BV dataset.

(lin) are very similar to each other (the drift seems almost linear in the windows) and decrease from $t = 430\,\mathrm{s}$ onward over a time period of the length of the time window, indicating a jump drift at $t = 430\,\mathrm{s}$.

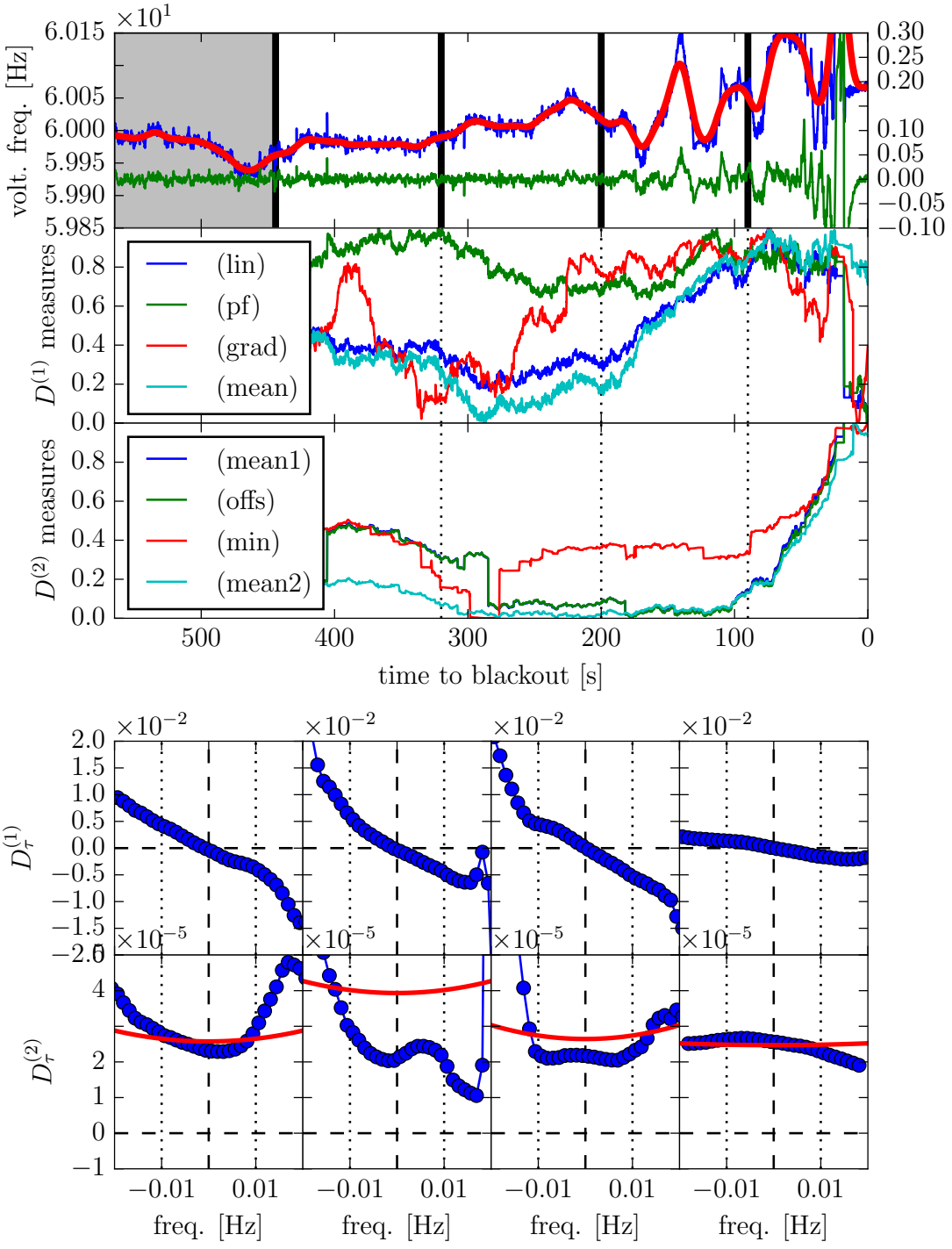


FIGURE 4.20: Drift & Diffusion estimation for the CD dataset.

The diffusion measures are all very similar to each other. They are almost constant until $t = 100\,\mathrm{s}$, at what point they start increasing.

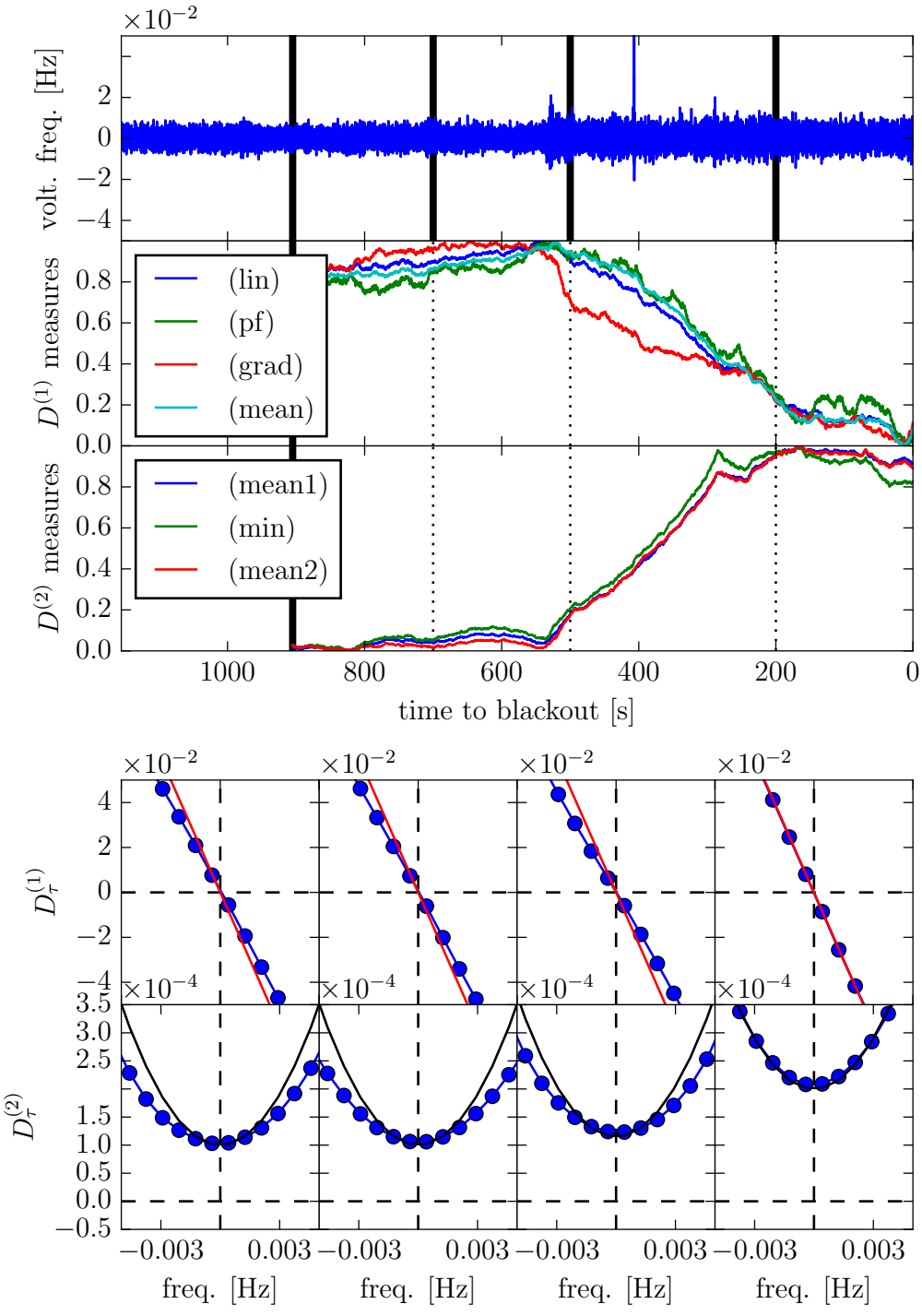


FIGURE 4.21: Drift & Diffusion estimation for the BV increment dataset.

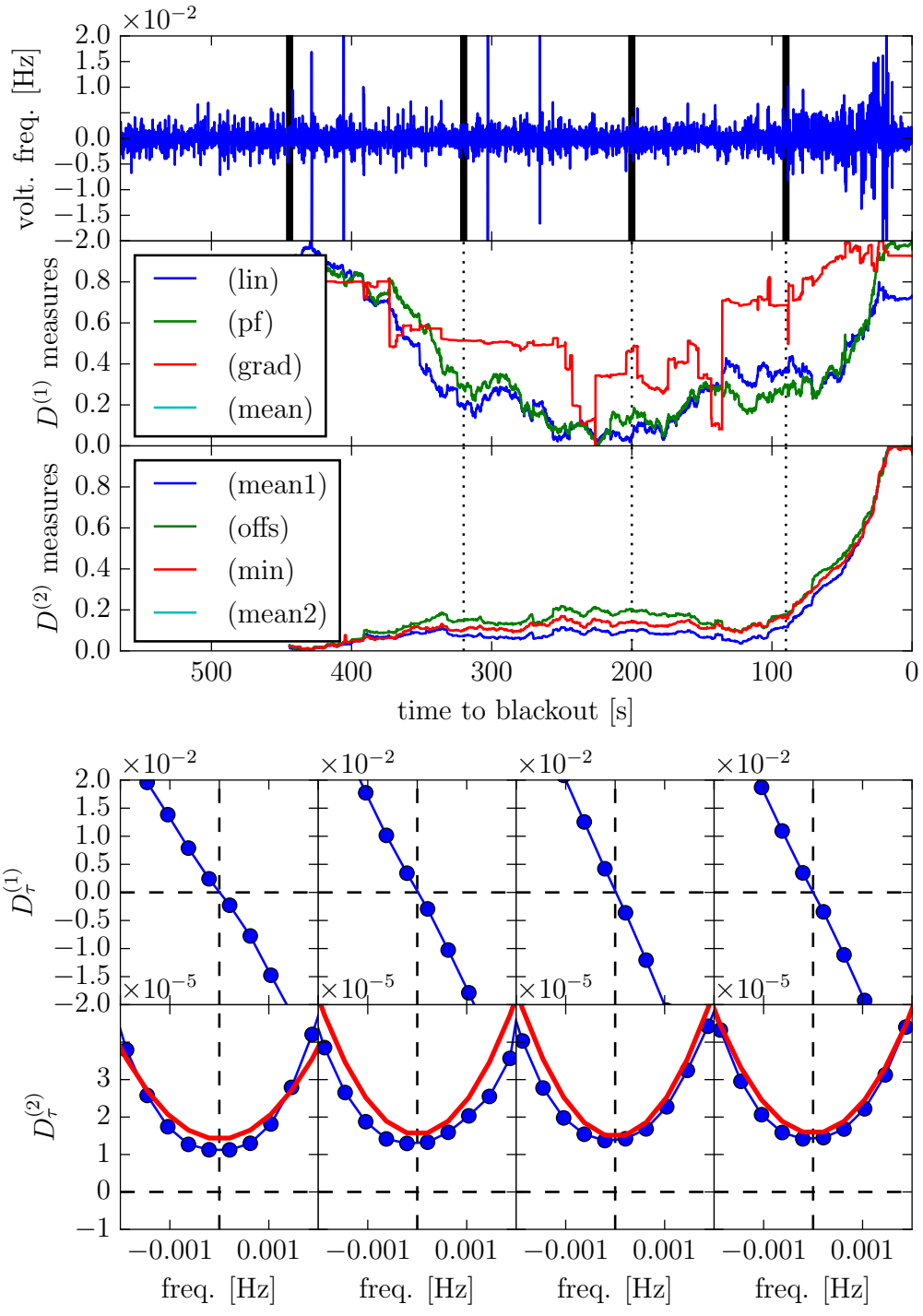


FIGURE 4.22: Drift & Diffusion estimation for the CD increment dataset.

To connect the concept of CSD to these data sets one can calculate some potential function by 'cumulative integration' of $D^{(1)}$ and setting its minimum to zero (see fig. 4.23). If CSD applies the potential must get flatter (broader) as time progresses.This can only be seen for the CD time series. For the BV (increment) time series and the CD increment time series the potentials gets slightly more steep, what contradicts the concept of CSD on these selected time windows.

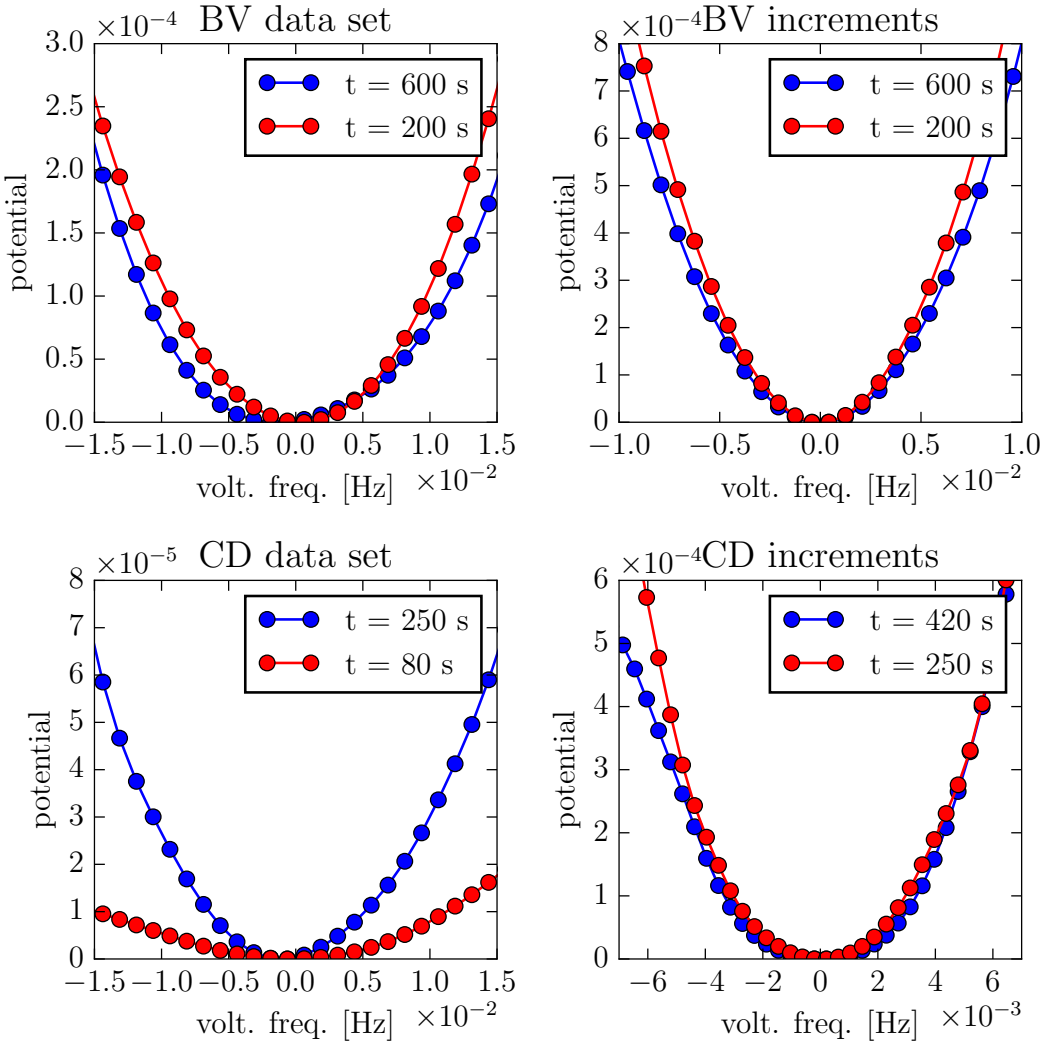


FIGURE 4.23: Potentials of each data set at distinctive points of time.

### 4.3.3 Markov time scale estimation

In this section we examine the Markov time scale as proposed in chapter 3.5. For this the BV and the CD dataset were again detrended with a Gaussian kernel filter as used above. The estimation of the pdfs was realized with a two-dimensional Nadaraya-Watson estimator with a bandwidth selection according to Silverman's rule of thumb and a Gaussian kernel. The time steps where taken as multiples of the sampling interval, as one can replace $\Delta t$ with the time lag in steps. The integral of equation 3.30 transitions into a matrix multiplication in this discrete case. As measure of distance the sum of squares was chosen. For both datasets the interval of comparison was taken as $[-0.01, 0.01]\,\mathrm{Hz}$ (blue vertical lines in figure 4.26) in the conditional variable and the complete data range for the non-conditional variable. For the calculation a grid of $200 \times 200$ points was used. An exemplary plot of the conditional pdfs is shown in figure 3.30, these get calculated for each time lag. The BV data set had its periodicity removed in fourier space, since the examination of the Markov property obviously is very susceptible to periodicity. The distant measure dependant of time lag can be seen in figure 4.24 for the CD data set and in figure 4.25 for the BV data set for four different time windows respectively. A shift of minimum as proposed in [4] could not be found.
Although a shift in minimum could not be found, an overall scaling of the distance measure dependent on the time shift was visible. The values of the CKe distance for a lag of 4 for the CD dataset and a lag of 2 can be seen in figures 4.27 and 4.28. This can be interpreted as a relative measure on how well the CKe is fulfilled along the time series. In both cases the course of this measure is very close to that of the slope of the drift estimations of the increment time series. The Markov property is getting fulfilled better as one approaches the outage. The drop-offs of value occur before the effect of csd.

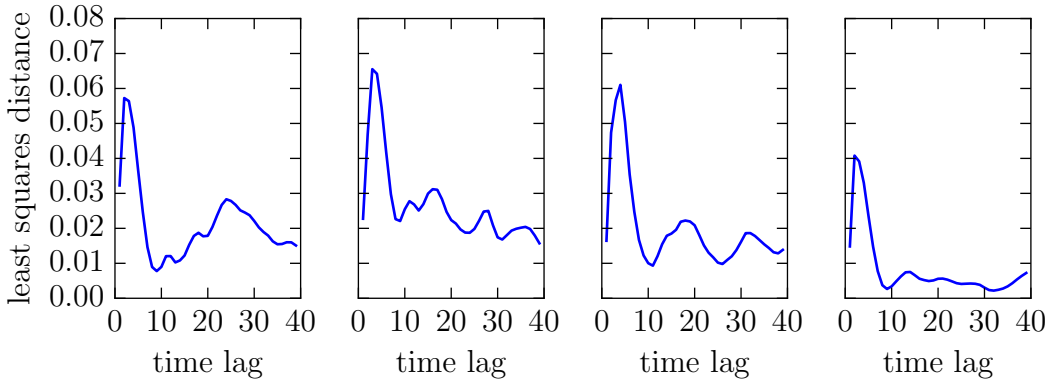


FIGURE 4.24: CD data set: CKe distance for windows of 5000 steps width at $t = 555\,s$, $t = 410\,s$, $t = 290\,s$ and $t = 170\,s$

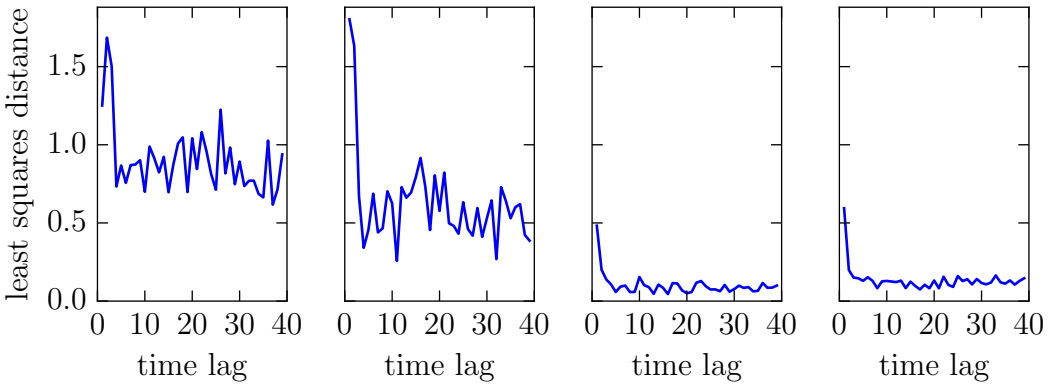


FIGURE 4.25: BV data set: CKe distance for windows of 5000 steps width at $t = 1155\,s$, $t = 855\,s$, $t = 555\,s$ and $t = 305\,s$

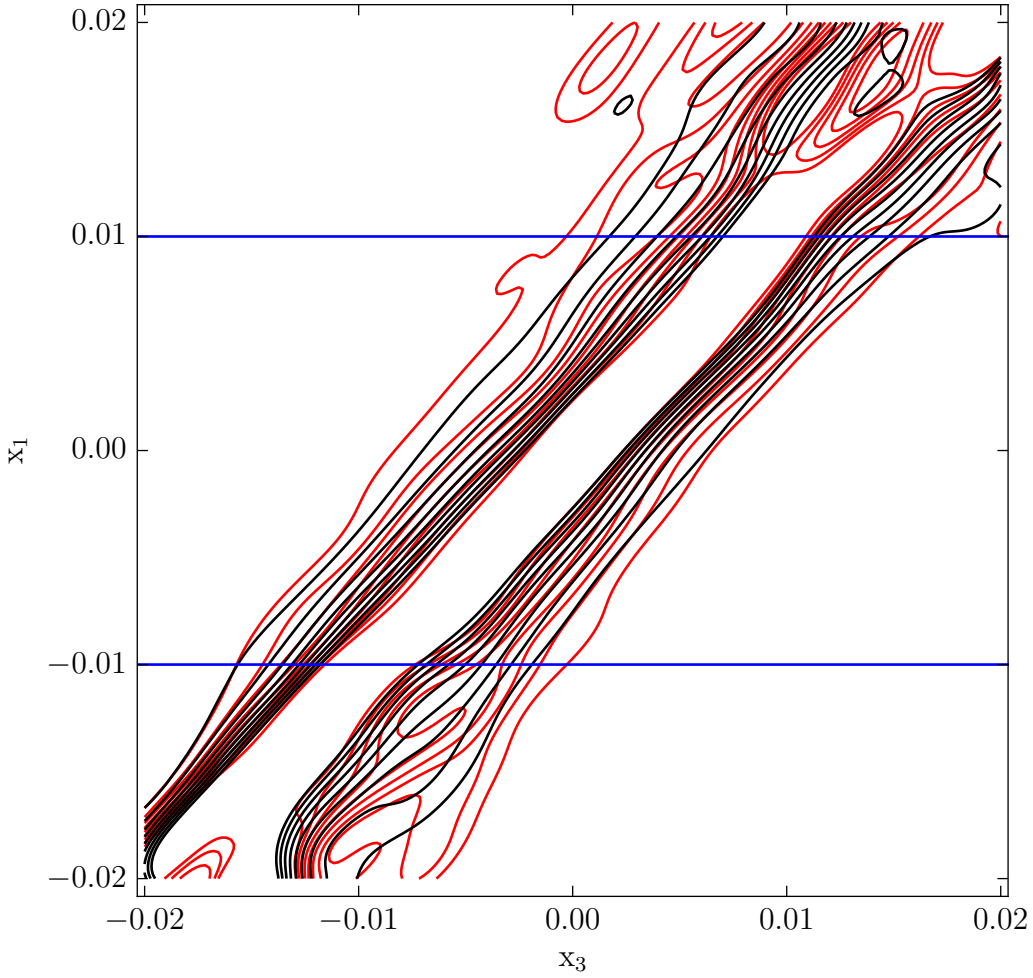


FIGURE 4.26: Contour plots of the conditional pdfs for lag-3 of the CD data set. Red: directly estimated $p(x_1|x_3)$. Black: Transition pdf calculated with CK equation.

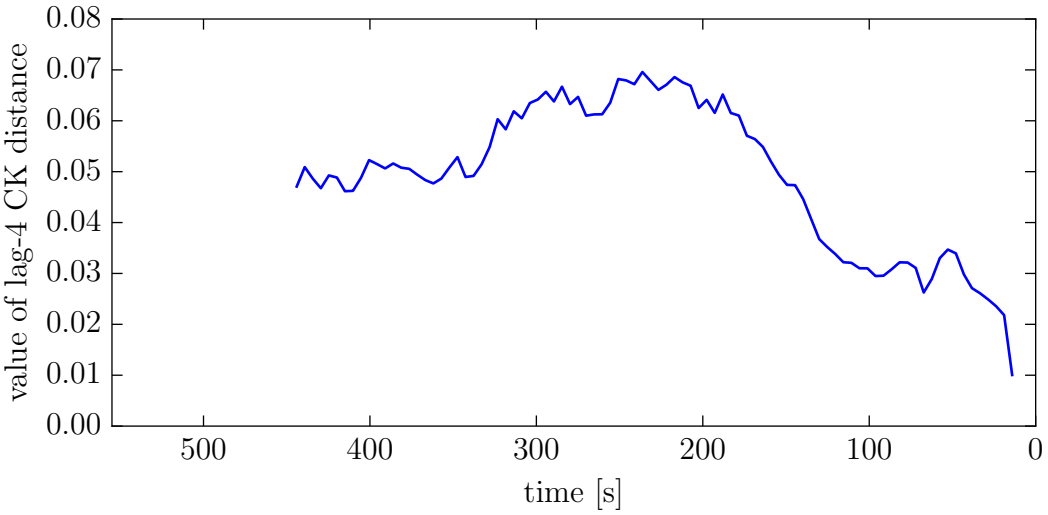


FIGURE 4.27: CKe distance for lag-4 of the CD data set.

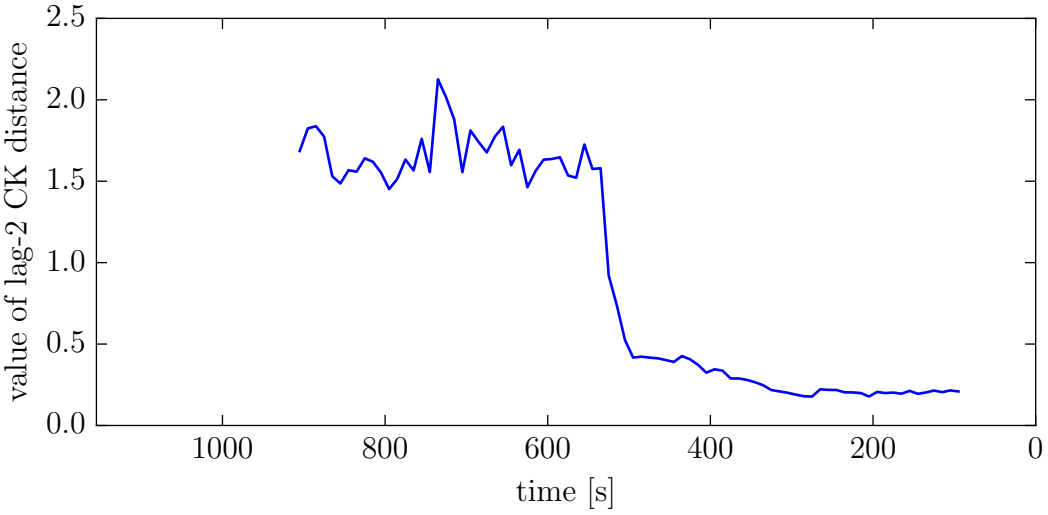


FIGURE 4.28: CKe distance for lag-2 of the BV data set.

# Chapter 5

# Discussion

The different measures exhibited varying usefulness, depending on what data set was examined. The variance as most obvious measure for 'critical slowing down' (csd) increased for the supercritical pitchfork bifurcation (spb) and the slow variable of the Haken-Zwanzig system (HZw) as it was expected. In the outage time series an increase of variance was only visible for the CD dataset around $150\,s$ before the outage, suggesting the occurrence of csd. In the BV dataset this was only visible very shortly before the outage.
For the CD dataset the variance of the increment time series increases at around the same point as the variance of the time series increases. This suggests additionally to the csd, the presence of a stronger fluctuating force. The BV increment time series of data set does not show this increase at the end of the time series, but at around $t = 530\,\mathrm{s}$ the variance increases spontaneously, indicating stronger fluctuations.
The skewness and kurtosis deemed themselves unsuitable because of the higher weightening of outliers (third and fourth power), especially for datasets with small sampling intervals.

An $a_1$-coefficient close to zero suggests uncorrelated values and close to one suggests a very long correlation time at lag-1. A negative value means an anti-correlation. As csd implies a wider spread of the dynamics a value closer to one is to be expected as a bifurcation is approached. For the spb this is certainly fulfilled; the value increases up to almost one as the bifurcation point gets crossed. An increase for the slow and the fast variable of the HZw system can be seen, suggesting a slowing of the dynamics, that is not connected to a rise of the variance for the fast variable. For the BV dataset the drop of value suggests a shortening of correlation time at this point, being the opposite of what was to be expected near a bifurcation. In contrast to that the $a_1$-coefficients for the CD data set approaches the value of one, exhibiting a behaviour as the spb, thus supporting the assumption of csd.
The $a_1$-coefficients for the increment time series of the BV data set are close to zero until $t = 530\,\mathrm{s}$, suggesting the Markov property on this scale of lag-1. After that point the increments get anti-correlated.
The two data sets exhibit contrary behaviour. Especially the measures for the BV data set at $t = 530\,\mathrm{s}$ suggest a moving away from the bifurcation point. The drift and diffusion estimation offers the possibility to examine the dynamics of such a time series. For the BV data set the drift gets stronger at $t = 530\,\mathrm{s}$, while

also the diffusion increases. This means the system is moving away from a bifurcation. For the increment time series the white noise limit is reached shortly after this point. For the CD data set the drift gets stronger from $t = 400\,\mathrm{s}$ to $t = 290\,\mathrm{s}$. After that, the drift approaches the bifurcation point and stays very close to it from $t = 80\,\mathrm{s}$ onward. At this point also the diffusion gets stronger. An explanation of this series of events can be best presented if one compares the measures for both data sets and connects them to the system disturbance report for this event. This is done in figure 5.1; the grey areas mark the used window sizes. Table 5.1 explains the measures shown in this figure. Table 5.2 shows events over the course of this power outage, marked by dotted lines in the figure.
The build-up of the outage lasted for longer than the time series shows. The first event (A) that occurs at some point within the time series is at $t = 717\,\mathrm{s}$ and no change of measures can be detected at this point. The changes of measures at $t = 530\,\mathrm{s}$ correspond to no event of the report. An explanation for this can be some safety mechanism setting in, that deliberately tightens the system. This was not detected by the variance, because the thightening of the drift and the rise of diffusion have reverse effects on the pdf (see equation 2.45 for the stationary Ornstein Uhlenbeck process) and cancel each other out.
The events (B)&(C) are very close to each other and are visible in both time series by an outlier. Event (C) is described in the report as the 'beginning of the disturbance'. At this event the drift of the CD data set gets stronger. This could be due to a response to this particular disturbance.
For both data sets the end of the times series exhibit the signs of csd.

Summarizing one can say, that the features of csd can be seen in both data sets close before the point of separation. At these sections of the time series the dynamics become obviously more volatile - and the increase of variance is a good measure to indicate this. However - this just happens very close to the outage, perhaps even crossing some point of no return. The more informative method of drift and diffusion estimation shows significant changes before the effect of csd. This is, that the dynamics tighten and the correlation time decreases. If this behaviour is due to some intentional control mechanisms or is a quality of such power systems remains unknown.
If one wants to apply a measure as an early-warning-signal, one would first establish some expected value of this measure for normal operation. Additionally one would establish some interval of error, on which this measure may fluctuate. Now, if the measure leaves this interval, one would use this as an indicator, that an outage is imminent. As measures for csd one can use variance as well as the $a_1$-estimation. As measures for the effect of tightening of the dynamics before csd, one can best use the drift estimation of the time series. This has been done for both time series and can be seen in figure 5.2.

TABLE 5.1: Explanation of the different measures displayed in figure 5.1

| label | measure | remarks about unscaled measures |
|---|---|---|
| 1 (green) | BV: pitchfork fit of drift | drives very close towards $\epsilon = 0$ at $t = 410\text{s}$ (bifurcation point). Afterwards, moves rapidly away from the bifurcation point |
| 2 (blue) | CD: $a_1$ of increments | |
| 3 (red) | CD: linear fit of drift of increments | drives towards, but not close the white noise limit from $t = 410\text{s}$ onward and returns to its starting value afterwards |
| 4 (yellow) | BV: $a_1$ | |
| 5 (green) | BV: $a_1$ of increments | |
| 6 (magenta) | BV: linear fit of drift of increments | slowly drives away from white noise limit until $t = 530\,s$, there strongly drives towards it and crosses it at $t = 470\,s$ |
| 7 (red) | BV: mean value of drift | moves away from bifurcation point at $t = 530\,\text{s}$ |
| 8 (blue) | BV: variance of increments | |
| 9 (cyan) | CD: mean value of drift | moves away from the bifurcation point until $t = 290\,\text{s}$, at that point it turns and moves closer towards the bifurcation point; staying very close to it at $t = 80\,\text{s}$ |
| 10 (blue) | BV: mean value of diffusion | increases until $t = 280\,\text{s}$, at what point it decreases again |
| 11 (green) | BV: mean value of increments of diffusion | moves towards white noise limit from $t = 530\,\text{s}$ |
| 12 (yellow) | CD: variance of increments | |
| 13 (red) | CD: mean value of diffusion | |
| 14 (magenta) | CD: variance | |
| 15 (cyan) | CD: mean value of diffusion of increments | white noise limit is reached as outage is imminant |

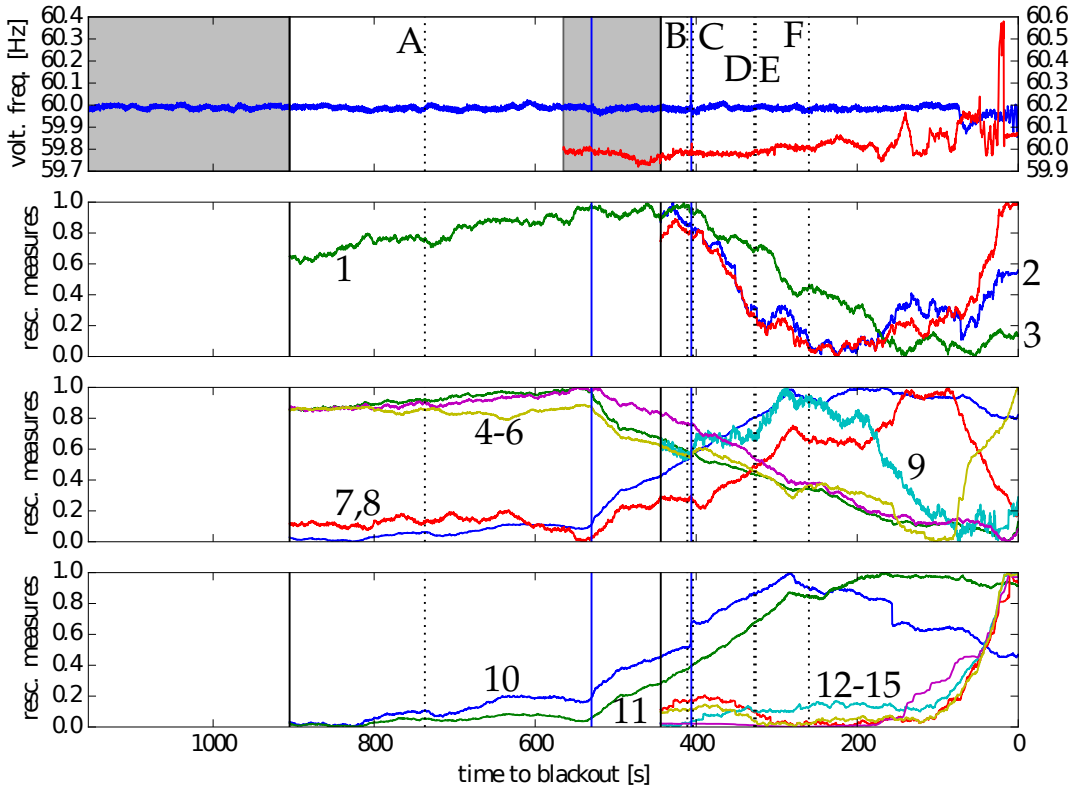


FIGURE 5.1: The different measures for both outage time series.

TABLE 5.2: Series of events of the 1996 power outage as in [29]

| event label | event |
|---|---|
| A | 500 kV Keeler-Allston single phase-to-ground fault; Flashed to tree |
| B | 115kV Merwin-St. Johns line opened due to relay misoperation. |
| C | 230kV Fault on Ross-Lexington line. Flashed to tree (**Beginning of disturbance**) |
| D | 500kV Captain Jack-Olinda line open at Captain Jack (**COI now open between Oregon and California**) |
| E | 500kV Midway-Vincent No. 3 line opens due to voltage collapse (**Northern California now separated from southern California**) |
| F | last recorded event |

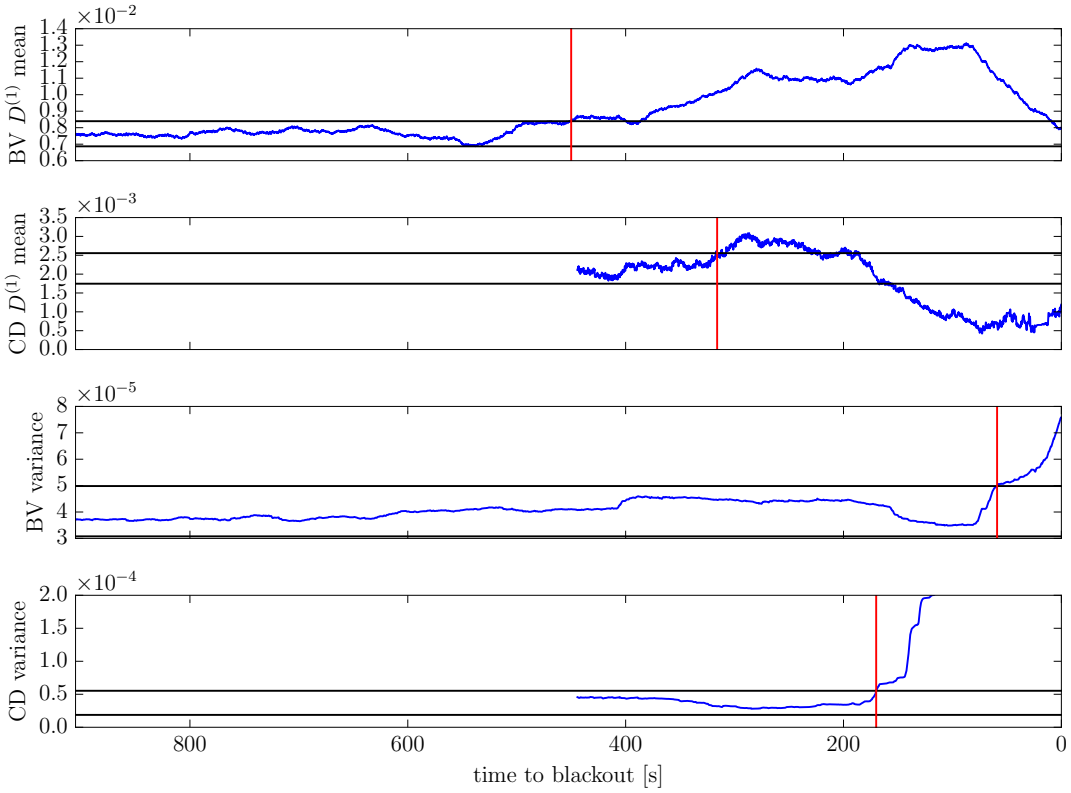


FIGURE 5.2: Early warning measures for power outage data. The intervals were chosen as multiples of the sample standard deviation ($\sigma_S$) around the sample mean of the first 10 percent of the data sets. These intervals are shown as black horizontal lines. The first crossing of these intervals are shown as red vertical lines. The upper two plots show the onset of tightening of the system with the mean drift value as measure for the BV ($10\sigma_S$ at $t_{\text{warning}} = 450\,\text{s}$) and CD ($6\sigma_S$ at $t_{\text{warning}} = 316\,\text{s}$) data set. The lower two plots show the onset of csd with variance as measure for the BV ($30\sigma_S$ at $t_{\text{warning}} = 63\,\text{s}$) and the CD ($30\sigma_S$ at $t_{\text{warning}} = 167\,\text{s}$) data set.

# Bibliography


[1] E. Cotilla-Sanchez, P.D.H. Hines, and C.M. Danforth. "Predicting Critical Transitions From Time Series Synchrophasor Data". In: *IEEE Transactions on Smart Grid* 3.4 (Dec. 2012), pp. 1832–1840.

[2] P. Hines, E. Cotilla-Sanchez, and S. Blumsack. "Topological Models and Critical Slowing down: Two Approaches to Power System Blackout Risk Analysis". In: *2011 44th Hawaii International Conference on System Sciences* (Jan. 2011), pp. 1–10.

[3] P. Manshour, J. Peinke, and M.R.R. Tabar. "Turbulencelike Behavior of Seismic Time Series". In: *Physical Review Letters* (2009).

[4] K. Kaviani and J. Peinke. "Short-Term Prediction of Medium and Large-Size Earthquakes Based on Markov and Extended Self-Similarity Analysis of Seismic Data". In: *Lecture Notes In Physics* (2006).

[5] M. Scheffer et al. "Early-warningsignals for critical transitions". In: *Nature* 461 (2009), pp. 53–59.

[6] V. Dakos, M. Scheffer, et al. "Slowing down as an early warning signal for abrupt climate change". In: *PNAS* 105.38 (Sept. 2008), pp. 14308–14312.

[7] V. Dose and A. Menzel. "Bayesian analysis of climate change impacts in phenology". In: *Global Change Biology* 10 (Feb. 2004), pp. 259–272.

[8] V. Dose and A. Menzel. "Bayesian correlation between temperature and blossom onset data". In: *Global Change Biology* 12.8 (2006), pp. 1451–1459.

[9] V.N. Livina and T.M. Vaz Martins. "Tipping point analysis of atmospheric oxygen concentration". In: *Chaos* 25 (Feb. 2015), pp. 1–12.

[10] R.P. Adams and D.J.C. MacKay. "Bayesian Online Changepoint Detection". In: *ARXIV* (Oct. 2007).

[11] A.S. Gsell et al. "Evaluating early-warning indicators of critical transitions in natural aquatic ecosystems". In: *Proceedings of the National Academy of Sciences* 113.50 (2016), E8089–E8095.

[12] T. Chadefaux. "Early warning signals for war in the news". In: *Journal of Peace Research* 51.1 (2014), pp. 5–18.

[13] M.G.M. Olde Rikkert et al. "Slowing Down of Recovery as Generic Risk Marker for Acute Severity Transitions in Chronic Diseases". In: *Critical Care Medicine* 44.3 (Mar. 2016), pp. 601–606.

[14] I.A. van de Leemput et al. "Critical slowing down as early warning for the onset and termination of depression". In: *PNAS* 111.1 (Jan. 2014), pp. 87–92.

[15] V. Dakos and S.R. Carpenter. "Methods for Detecting Early Warnings of Critical Transitions in Time Series Illustrated Using Simulated Ecological Data". In: *PLoS ONE* (2012).

[16] P. Rinn et al. "Stochastic method for in-situ damage analysis". In: *The European Physical Journal B* 86.1 (2012), p. 3.

[17] S.R. Carpenter and W.A. Brock. "Stochastic method for in-situ damage analysis". In: *Ecology* 92.12 (2011), pp. 2196–2201.

[18] M. Scheffer et al. "Anticipating Critical Transitions". In: *Science* 338 (Oct. 2012), pp. 344–348.

[19] H. Risken. *The Fokker-Planck Equation*. Springer, 1996.

[20] W. Paul and J. Baschnagel. *Stochastic Processes*. Springer, 1999.

[21] C.W. Gardiner. *Handbook of Stochastic Methods*. Springer, 1983.

[22] N.G. van Kampen. *Stochastic Processes in Physics and Chemestry*. NH PL, 1992.

[23] H. Haken. *Synergetik*. Springer, 1982.

[24] S.H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*. Westview Press, 2014.

[25] H. Rinne. *Taschenbuch der Statistik, 4. Auflage*. Verlag Harri Deutsch, 2008.

[26] U. Hassler. *Stochastische Integration und Zeitreihenmodellierung*. Springer, 2007.

[27] C. Honisch and R. Friedrich. "Estimation of Kramers-Moyal coefficients at low sampling rates". In: *Physical Review* 83 (June 2011), pp. 1–6.

[28] D. Lamouroux and K. Lehnertz. "Kernel-based regression of drift and diffusion coefficients of stochastic processes". In: *Physics Letters A* 373 (July 2009), pp. 3507–3512.

[29] North American Electric Reliability Council. *Review of Selected 1996 Electric System Disturbances in North America*. Aug. 2002. URL: `http://www.nerc.com/pa/rrm/ea/System%20Disturbance%20Reports%20DL/1996SystemDisturbance.pdf`.

# *Danksagung*

An dieser Stelle möchte ich mich bei allen bedanken, die mich während des Erstellens dieser Masterarbeit unterstützt haben.

Ganz besonders herzlich möchte ich mich bei meinem Betreuer Dr. Oliver Kamps bedanken, der mir ein spannendes Thema geboten und immer mit großer Gefälligkeit Zeit für mich gefunden hat.

Sehr herzlichen Dank auch an Prof. Uwe Thiele für den Platz in der Arbeitsgruppe und seine Bemühungen zur Datenbeschaffung.

Ausserordentlicher Dank gilt meinen Eltern Angelika und Michael Ehebrecht, welche mir das Studium gerne ermöglicht und mich immer mit großem Interesse unterstützt haben. Ich danke vor allem für den ausdauernden Beistand während der Operationen.

Frank Ehebrecht
Münster, 19.03.2017

## Plagiatserklärung der / des Studierenden

Hiermit versichere ich, dass die vorliegende Arbeit über _____ _____ selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

_____
(Datum, Unterschrift)

Ich erkläre mich mit einem Abgleich der Arbeit mit anderen Texten zwecks Auffindung von Übereinstimmungen sowie mit einer zu diesem Zweck vorzunehmenden Speicherung der Arbeit in einer Datenbank einverstanden.

_____
(Datum, Unterschrift)