
**Modelling the variability of complex
systems by means of Langevin
processes**

**On the application of a dynamical approach
to experimental data**

Julia Gottschall

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines

**DOKTORS DER NATURWISSENSCHAFTEN
DR. RER. NAT.**

angenommene Dissertation

von Frau Julia Gottschall
geboren am 14.03.1982 in Göttingen

Gutachter: Prof. Dr. Joachim Peinke
Zweitgutachterin: Prof. Dr. Ulrike Feudel
Tag der Disputation: 20.01.2009

Contents

Abstract – Zusammenfassung	vii
1 Introduction	1
1.1 Modelling variability of complex systems	1
1.2 Reconstructing the dynamics of Langevin processes	3
1.3 Stochastic modelling of experimental data	5
1.4 Extension to Langevin-like processes	7
1.5 Outline of this thesis	9
Bibliography	10
2 Definition and handling of different drift and diffusion estimates	13
2.1 Introduction	13
2.2 Finite sampling rates	16
2.3 Presence of measurement noise	19
2.4 Combination of low sampling rates and measurement noise	22
2.5 Reconstruction through optimization	25
2.6 Handling the estimates and a note on the robustness of fixed points	28
2.7 Conclusions	30
2.8 Appendix: Derivation of third-order terms	31
Bibliography	32
3 A phenomenological power performance model	35
3.1 Basic considerations on power performance testing	36
3.1.1 Power performance and effective control	36
3.1.2 Standard procedure for power performance testing (IEC 61400-12-1)	37
3.1.3 Averaged power performance versus short-term dynamics	39
3.2 Turbulent structures in wind speed and power output	41
3.2.1 Characterization of short-term structures in wind speed time series	41
3.2.2 A note on gusts	47
3.2.3 Transfer of turbulent structures to the power output	50
3.3 Definition and reconstruction of a dynamical power characteristic	55
3.3.1 Stochastic relaxation model	55
3.3.2 Reconstruction of dynamical power characteristic	57
3.4 Discussion of the model	62
3.4.1 Self-consistency	62
3.4.2 Time scale of description	65
3.4.3 Representativity of considered wind speed	67
3.4.4 Relevance of the dynamical power characteristic	70
3.5 General potential of a stochastic fixed-point analysis	72
3.6 Appendix A (Publication in Environ. Res. Lett.)	75

3.7 Appendix B: Data description	84
Bibliography	87
4 Stochastic modelling of human postural control	91
4.1 Exploring the dynamics of balance data	92
4.1.1 Introduction	92
4.1.2 Experimental setup and data analysis	93
4.1.3 Results	99
4.1.4 Conclusions	101
4.2 Distinguishing between task and time effects	103
4.2.1 Outline of the study	103
4.2.2 Results and discussion	104
4.3 Appendix: One-way ANOVA	106
Bibliography	107
5 Summary	111
Danksagung	115
Lebenslauf	117
Erklärung	119
Publikationsliste	120

List of Figures

2.1	Time series of an exemplary stochastic process	15
2.2	$D_{E,\tau}^{(1)}(x, \tau_0)$ for a process with cubic drift and constant diffusion . . .	18
2.3	The estimate $D_{E,\tau}^{(1)}(x, \tau_0)$	19
2.4	The estimate $D_{E,\sigma}^{(1)}(y)$	21
2.5	Empirical and theoretical estimates $D_{E,\tau}^{(n)}$ and $D_{E,\sigma}^{(n)}$ ($n = 1, 2$)	23
2.6	The estimate $D_{E,\tau\sigma}^{(1)}(y, \tau_0)$	24
2.7	Convergence behaviour of the estimates $D_{E,\tau}^{(1)}$ and $D_{E,\tau\sigma}^{(1)}$	25
2.8	Empirical and theoretical estimates $D_{E,\tau\sigma}^{(n)}$ and extrapolated offsets .	27
2.9	Theoretical drift estimates for four exemplary processes	29
3.1	Typical time series for $u(t)$, fluctuations $u'(t)$ and increments $u_\tau(t)$.	42
3.2	Turbulence intensity, skewness and kurtosis for $u(t)$ and $u'(t)$, resp. .	43
3.3	Pdfs of high-frequency data u conditioned on 10 min periods	44
3.4	Entire pdfs of the fluctuations u' with respect to different values of T	45
3.5	Pdfs of the wind speed increments u_τ for different time increments τ	46
3.6	Standard deviation σ and form parameter λ^2 for pdfs of u_τ versus τ	47
3.7	Entire pdf $p(u_\tau)$ and conditioned pdfs $p(u_\tau \bar{u})$	48
3.8	Pdfs of the wind directions increments ϕ_τ for different τ	49
3.9	Pdfs for the vertical wind speed component u_z and its increments $u_{z,\tau}$	50
3.10	Turbulence intensity, skewness and kurtosis for $P(t)$ and $P'(t)$, resp.	51
3.11	Pdfs of high-frequency data P conditioned on 10 min periods	52
3.12	Pdfs of the power increments P_τ for different τ	53
3.13	Pdfs for the increments P_τ and $P_{p.c.,\tau}$ for $\tau = 20$ s	53
3.14	σ_P and λ^2 for the pdfs of P_τ and $P_{p.c.,\tau}$ versus τ	54
3.15	Illustration of relaxation model for power performance	56
3.16	Reconstructed power curves for numerical data with/without relaxation	57
3.17	Estimation of the drift coefficient from first conditional moment . . .	59
3.18	Illustration of fixed-point analysis	60
3.19	Dynamical power characteristic for two different power binnings . . .	61
3.20	Reconstructed fixed points for numerical power output data	63
3.21	Pdfs of reconstr. noise processes $\Gamma_i(t)$ for different wind speed bins .	64
3.22	Pdfs of empirical and reconstructed increments $P_\tau(t)$ for different τ .	65
3.23	Dynamical power characteristic for averaged wind speed \tilde{u}_T	66
3.24	Correlation between wind speed and power output time series	68
3.25	Reconstructed drift coefficients with/without additional delay	69
3.26	Reconstructed dynamical power curves with/without additional delay	69
3.27	Power characteristics for a turbine with multistable dynamics	72
3.28	Convergence behaviour of mean values and fixed points	74
3.29	Reconstructed power curves for simulated data for different values of I	78
3.30	Simulated data with non-ideal noise	80

3.31	Reconstructed power curves for simulated data with non-ideal noise	80
3.32	Schematic comparison of different types of power curves	81
3.33	Histograms of power output data for different bins	82
3.34	Site layout and location of the meteorological mast	85
3.35	Distribution of wind directions for the considered site	86
4.1	Experimental setup (balance experiment)	94
4.2	Angular displacement and velocity in medial-lateral direction	95
4.3	Reconstruction of drift and diffusion coefficient with binning	96
4.4	Direct reconstruction of drift and diffusion coefficient without binning	97
4.5	Self-consistency test – empirical and reconstructed pdfs	98
4.6	Performance of balance with and without supra-postural tasks	101
4.7	Experimental setup for sitting condition	103

Abstract

Central topic of this thesis is the modelling of complex systems as stochastic processes. In this regard, the underlying dynamics of the considered process is described by means of effective Langevin equations. The functions for drift and diffusion determining these equations are derived directly from a set of measured data, based on a procedure proposed by Friedrich and Peinke.

The focus of this thesis is on the application of this procedure to experimental data. In particular, respective disturbances are discussed that complicate a reliable estimation of the drift and diffusion coefficients. After an introduction to the theory forming the basis of the considered method and a critical overview over previous examples of application, two specific applications are discussed in more detail.

The first application is the modelling of a wind turbine's power output on the basis of measured wind speed and power data. After a detailed characterization of the small-scale fluctuations of atmospheric wind speed that act on the turbine, a model for the transfer of these fluctuations to the power output is introduced and discussed. The dynamical relaxation behaviour of the turbine is described in terms of a set of one-dimensional Langevin equations, and the governing coefficients are respectively reconstructed. The derived drift coefficients and, in particular, the fixed points of the deterministic part of the dynamics reconstructed in this way are utilized to define a dynamical power characteristic for a specific type of wind turbine.

In a second application, the process of human postural control is analyzed with respect to a balance experiment. Measured variations in angular velocity of a balance board on that the test subjects balance are modelled as Langevin process. It is supposed that the reconstructed drift part reflects an internal control behaviour, similar to the example of the wind turbine dynamics. The impact of so-called supra-postural tasks is investigated by means of a differentiated evaluation of drift and diffusion coefficients.

Zusammenfassung

Zentrales Thema dieser Arbeit ist die Modellierung komplexer Systeme im Sinne von stochastischen Prozessen. Die jeweils zugrunde liegende Dynamik wird dabei durch effektive Langevin-Gleichungen beschrieben, und die diese Gleichungen bestimmenden Funktionen für Drift und Diffusion werden auf der Basis eines von Friedrich und Peinke entwickelten Verfahrens direkt aus einem gemessenen Datensatz bestimmt.

Der Schwerpunkt der vorliegenden Arbeit liegt dabei auf der Anwendung der Methode auf experimentelle Daten. Insbesondere werden entsprechende Beeinträchtigungen diskutiert, die eine zuverlässige Schätzung von Drift- und Diffusionskoeffizienten verhindern. Je nach Beeinträchtigung werden verschiedene Schätzwerte definiert und deren Abweichungen zu den intrinsischen Funktionen untersucht. Nach einer Einführung in die dem Verfahren zugrunde liegende Theorie und einem kritischen Überblick über bisher publizierte Anwendungsbeispiele werden zwei Anwendungen näher untersucht.

Die erste Anwendung ist die Modellierung der Leistungsabgabe einer Windenergieanlage auf der Basis von gemessenen Windgeschwindigkeits- und Leistungsdaten. Nach einer detaillierten Charakterisierung der auf die Anlage einwirkenden kleinskaligen atmosphärischen Windfluktuationen wird ein Modell für die Übertragung dieser Fluktuationen auf die Leistungsabgabe der Windenergieanlage eingeführt und diskutiert. Das dynamische Relaxationsverhalten der Anlage wird dabei durch ein System eindimensionaler Langevin-Gleichungen beschrieben und entsprechend rekonstruiert. Die berechneten Driftkoeffizienten und insbesondere die Fixpunkte des so ermittelten deterministischen Anteils der Dynamik werden benutzt, um eine dynamische Leistungscharakteristik für einen speziellen Anlagentyp zu definieren.

In einer zweiten Anwendung wird das menschliche Gleichgewichtsverhalten in Rahmen eines Balancierexperiments analysiert. Die gemessenen Änderungen der Winkelgeschwindigkeit eines Kreisels, auf dem die Testpersonen balancieren, werden wiederum als Langevin-Prozess modelliert. Dabei wird angenommen, dass der rekonstruierte Driftanteil, ähnlich wie im Beispiel der Windenergieanlage, das interne Regelungsverhalten wiedergibt. Der Einfluss sogenannter supra-postural tasks, übergeordneter Aufgaben, die die Testpersonen auszuführen haben, wird durch eine differenzierte Auswertung von Drift- und Diffusionskoeffizienten untersucht.

Chapter 1

Introduction

1.1 Modelling variability of complex systems

Variability is such a general term that it might call up quite diverse associations. For this thesis, I want to define variability by means of fluctuations or fluctuating time series, respectively, relating the abstract term to a measurable quantity. But also the explicit fluctuations may be defined in various different manners, depending on the specific field of application and the more detailed framework of the respective analysis. In this section, the range of issues from measuring fluctuations of a certain process to an appropriate modelling of the respective underlying dynamics is covered. The term dynamics is, in this sense, related to the internal forces that determine the characteristics of a certain system, and reflects something like an intrinsic control behaviour in contrast to external disturbances.

To start with, the output of a measurement is often given by a fluctuating time series $x(t)$, where x is the observed signal and t denotes the discretization of time. A measurement may refer to a kind of physical sampling procedure just as well as any other kind of quantitative investigation. And the signal may be the position of a particle but also some socio-economic data, where the units of t range from the fraction of a second to hours, days or months. The fluctuations of the values x reflect the variability of the investigated process and eventually also of the underlying system. This variability may be expressed by the variance or standard deviation of the observed time series by applying standard sample statistics or in terms of the (auto)correlation function of the data detecting possible finite correlations in time. A still more detailed description of the fluctuations is obtained by revealing the dynamical equation that underlies the observed process. The basis of such a dynamical approach is the formulation of a differential equation, e.g. in terms of an ordinary differential equation (ODE) $\dot{\xi}(t) = f(\xi(t))$, which is called a dynamical system or the effective dynamical equation of the process. $\xi(t)$ denotes the state variable of the process that takes the value $x(t)$ when measured at time t . The right side of the equation summarizes all the forces that act or arise within the system and therewith explains the evolution of the quantity ξ in time on the left side. The simple first-order ODE given above may be generalized to an ODE of order n including the n -th derivative of the variable x , that in turn can be expressed as n -dimensional set of first-order ODEs. A simulation, i.e. an integration of the dynamical equation gives a time series $x(t)$. The other way round, the dynamics of the variable ξ , i.e. $\dot{\xi}(t)$, can be in principle just as well reconstructed directly from the observed time series even though this is in reality not always straightforward and requires specific reconstruction schemes.

The latter is the objective of the field of time series analysis. This discipline

covers methods for the characterization of an observed time series, the forecast of its further evolution as well as the monitoring of possible systematic changes in the corresponding dynamics. A prominent part of it is the study of nonlinear effects. By the influence of nonlinear terms in the dynamical equations its solutions behave not any longer in an additive way. A corresponding time series may show irregular fluctuations though still being governed by deterministic dynamics. A sensitive dependence on initial conditions that is associated with this kind of behaviour is often called deterministic chaos. For an overview of the most common methods in nonlinear time series analysis it is referred to [Kantz & Schreiber 1997].

Nonlinear approaches are of particular importance for the study of complex systems. A complex system is a system that is composed of many different subsystems but shows a behaviour that cannot be described solely by the dynamics of the single building blocks in a sufficient way. A typical property of complex systems is the process of self-organization, i.e. the autonomous formation of certain patterns and structures within the system. Amongst others, complex systems are e.g. found in the field of biomechanics. To give a simple example – we are able to describe the movement of a single particle that is exposed to well-defined external forces by a quite simple dynamical equation, but it is considerably more difficult to capture the influence of human coordination and control which can be considered as a complex structure. Complex structures are formed as the result of very complicated, nonlinear and often only insufficiently known interactions.

The probably most adequate description of complex systems is given by Haken's synergetics approach [Haken 1983]. Haken found that the dynamics of a wide range of complex systems takes place on different time scales that can be separated adiabatically. In this case, the degrees of freedom of the system can be reduced to a small number of order parameters whose evolution is described by stochastic differential equations. The order parameters dominate the macroscopic dynamics of the system while the microscopic degrees of freedom appear only as fluctuating forces, i.e. a kind of noise, in the dynamical equation.

Haken's explicit considerations about different time scales motivate the introduction of stochastic processes for the modelling of the fluctuations of a complex system. The formulation of a dynamical equation that incorporates deterministic as well as stochastic terms corresponds to a separation of internal and external forces for an open system. The stochastic terms sum up the external forces without giving a detailed description of the single components. In this brief introduction, I do not want to go into detail about the question if this stochasticity is a necessary component or rather a rough simplification. This definitely depends on the specific application, the actual complexity of the considered system, as well as on the objective of the respective procedure of data analysis.

In order to describe stochastic aspects, a couple of different theoretical concepts can be considered. In this thesis, however, I focus on the formulation of stochastic processes in terms of diffusion equations given by stochastic first-order differential equations in time. For the variable $\xi(t)$ defining the state of the dynamical system, let $h(\xi, t)$ denote a force acting on the system at time t when $\xi(t)$ equals x . Considering a second function g that acts together with a noise term $\Gamma(t)$ as stochastic force and assuming $\Gamma(t)$ to define Gaussian distributed δ -correlated white noise, the stochastic evolution equation of $\xi(t)$ is given by the so-called Langevin equation

$$\frac{d}{dt}\xi(t) = h(\xi, t) + g(\xi, t) \Gamma(t) . \quad (1.1)$$

For this class of processes it is known how to reconstruct the functions h and g directly from measured data without any further knowledge about the internal dynamics of the considered system. This procedure is called stochastic modelling. The fundamental mathematical concept has been formulated in [Kolmogorov 1931].

Its application, i.e. a respective method of reconstruction, has been introduced in [Siegert et al 1998]. During the last years, this scheme has been applied to a large variety of problems and disciplines, among them many processes that are considered to stem from complex systems. In the following sections of this introduction, I firstly give a short outline of the theory of Langevin processes and its reconstruction before I discuss the central issues for an appropriate application of the method to experimental data and how the Langevin model is extended to more general processes.

1.2 Reconstructing the dynamics of Langevin processes ¹

For a multidimensional stochastic state variable $\boldsymbol{\xi}$ the general Langevin equations have the form

$$\dot{\xi}_i = h_i(\boldsymbol{\xi}, t) + \sum_j g_{ij}(\boldsymbol{\xi}, t)\Gamma_j(t) \quad (1.2)$$

for each component i . The functions h_i and g_{ij} may be arbitrary nonlinear functions of the components ξ_i and the time t . For constant g_{ij} , (1.2) is called Langevin equation with additive noise force, and accordingly with a multiplicative noise force for the functions g_{ij} depending on $\boldsymbol{\xi}$. The Langevin force $\boldsymbol{\Gamma}(t)$ corresponds to Gaussian distributed and δ -correlated stochastic variables with vanishing mean and a variance that equals 2, i.e.

$$\langle \boldsymbol{\Gamma}(t) \rangle = 0, \quad \langle \Gamma_i(t)\Gamma_j(t') \rangle = 2\delta_{ij}\delta(t-t'). \quad (1.3)$$

The δ -correlation of $\boldsymbol{\Gamma}(t)$ provides that Langevin processes are Markov processes, i.e. the future state of the system does only depend on its present state but not on the states in the past. For Markov processes the multiple conditioned probabilities reduce to simple conditioned probabilities according to

$$p(\mathbf{x}_{i+1}, t_{i+1} | \mathbf{x}_i, t_i; \dots; \mathbf{x}_0, t_0) = p(\mathbf{x}_{i+1}, t_{i+1} | \mathbf{x}_i, t_i), \quad (1.4)$$

and the joint probability distribution $p(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{x}_0, t_0)$ can be constructed from the knowledge of the conditional probability distributions $p(\mathbf{x}_{i+1}, t_{i+1} | \mathbf{x}_i, t_i)$.

Writing down the differential equation (1.2), this equation is expected to be integrable. But due to the properties (1.3) of $\boldsymbol{\Gamma}(t)$ it does not obey ordinary differential calculus. A common physical interpretation is to replace the δ -function by a function with very small finite width. The more mathematical way to obtain a consistent interpretation is to write (1.2) as integral equation according to

$$\xi(t+\tau) = \xi(t) + \int_t^{t+\tau} dt' h(\boldsymbol{\xi}, t') + \int_t^{t+\tau} dt' g(\boldsymbol{\xi}, t')\Gamma_j(t') \quad (1.5)$$

and introduce the relation

$$dW_j(t) \equiv W_j(t+dt) - W_j(t) = \Gamma_j(t)dt \quad (1.6)$$

where $W(t)$ is a Wiener process. The main difficulty is now to define the integral $\int G(t')dW(t')$ for an arbitrary function $G(t)$. There are two common definitions in the literature, the one according to Itô and that by Stratonovich. For this thesis I restrict my considerations to the interpretation by Itô that especially shows advantages for the numerical integration of the stochastic differential equations. In

¹The reasoning of this section basically follows the proceeding of [Risken 1989], and also conventions that may differ in other references are taken over from here.

this case and by the use of a simple Euler scheme for the integration of the functions h_i , (1.2) reads in a discrete formulation

$$\xi_i(t + \tau) = \xi_i(t) + \tau h_i(\boldsymbol{\xi}(t), t) + \sum_j \sqrt{\tau} g_{ij}(\boldsymbol{\xi}(t), t) \Gamma_j(t). \quad (1.7)$$

To exhibit a scheme for the reconstruction of the functions h_i and g_{ij} from a set of experimental data, the connection between the Langevin equations for the trajectories of a stochastic process and the corresponding Fokker-Planck equation for the statistical properties of these trajectories is shown. Following [Risken 1989], as first step the conditional moments

$$M^{(n)}(\mathbf{x}, t, \tau) := \langle [\boldsymbol{\xi}(t + \tau) - \boldsymbol{\xi}(t)]^n | \boldsymbol{\xi}(t) = \mathbf{x} \rangle \quad (1.8)$$

$$= \int d\mathbf{x}' [\mathbf{x}' - \mathbf{x}]^n p(\mathbf{x}', t + \tau | \mathbf{x}, t) \quad (1.9)$$

are introduced. The products $[...]^n$ refer to a dyadic multiplication, i.e. the objects $M^{(n)}(\mathbf{x}, t, \tau)$ may be vectors, matrices or tensors, dependent on the order n and the dimensionality of \mathbf{x} or $\boldsymbol{\xi}$, respectively. The conditioning in (1.8) means that only the values that fulfil $\boldsymbol{\xi}(t) = \mathbf{x}$ are considered for the calculation of the average. On the basis of the conditional moments, as a second step, the Kramers-Moyal coefficients are defined according to

$$D^{(n)}(\mathbf{x}, t) := \frac{1}{n!} \lim_{\tau \rightarrow 0} \frac{1}{\tau} M^{(n)}(\mathbf{x}, t, \tau). \quad (1.10)$$

The convergence of the limits is presumed here. By means of the Kramers-Moyal coefficients, the time evolution of a probability distribution function (pdf) $p(\mathbf{x}, t)$ can be described by formulating the partial differential equation

$$\frac{\partial}{\partial t} p(\mathbf{x}, t) = \left\{ \sum_{n=1}^{\infty} (-\nabla)^n D^{(n)}(\mathbf{x}, t) \right\} p(\mathbf{x}, t). \quad (1.11)$$

For a general process this expansion may consist of an infinite number of terms. But for a process that follows a Langevin equation this so-called Kramers-Moyal expansion is truncated after the second term on the right side and reduced to the Fokker-Planck equation

$$\frac{\partial}{\partial t} p(\mathbf{x}, t) = \left\{ - \sum_i \frac{\partial}{\partial x_i} D_i^{(1)}(\mathbf{x}, t) + \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\mathbf{x}, t) \right\} p(\mathbf{x}, t). \quad (1.12)$$

Note that the Fokker-Planck equation can be formulated in an analogous way for the conditional probability $p(\mathbf{x}', t + \tau | \mathbf{x}, t)$.

Following again [Risken 1989], one directly obtains by utilizing the properties of $\Gamma(t)$ the relations

$$D_i^{(1)}(\mathbf{x}, t) = h_i(\mathbf{x}, t), \quad D_{ij}^{(2)}(\mathbf{x}, t) = \sum_k g_{ik}(\mathbf{x}, t) g_{jk}(\mathbf{x}, t), \quad (1.13)$$

$$D^{(n)}(\mathbf{x}, t) = 0 \quad \forall n \geq 3. \quad (1.14)$$

In several references, it is also referred to the theorem of Pawula at this point. According to Pawula's theorem all coefficients $D^{(n)}(\mathbf{x}, t)$ for $n \geq 3$ vanish if $D^{(n=4)}(\mathbf{x}, t) = 0$ holds. The latter condition is fulfilled if the noise included in the considered process is Gaussian distributed – which is just one of the basic requirements for the noise source $\Gamma(t)$ in (1.2) and the respective dynamics to follow a Langevin process.

For a Langevin process, the functions $D^{(n)}(\mathbf{x}, t)$ ($n = 1, 2$) are referred to as drift and diffusion coefficients, respectively, and can be interpreted in the following way. For a trajectory that passes the point \mathbf{x} at time t , $D^{(1)}(\mathbf{x}, t)$ gives the average trend of movement and $D^{(2)}(\mathbf{x}, t)$ the corresponding variance. I.e. the distribution of the points $\mathbf{x}(t + \tau)$ for all t with $\boldsymbol{\xi}(t) = \mathbf{x}$, which the trajectory passes after the time increment τ , is given by a Gaussian distribution with mean $\mathbf{x}(t) + \tau D^{(1)}(\mathbf{x}, t)$ and variance $2\tau D^{(2)}(\mathbf{x}, t)$.

In principle, it is now straightforward to reconstruct drift and diffusion coefficients, and therewith an effective Langevin equation, for a set of experimental data. One simply has to calculate the conditional moments and subsequently the Kramers-Moyal coefficients according to (1.8–1.10). How this is realized for different applications and especially which difficulties may arise is presented in the next section. To simplify matters, I no longer make a difference between the stochastic state variable $\boldsymbol{\xi}$ and its value \mathbf{x} from now on, and write (1.2) equivalently as

$$\dot{x}_i = D_i^{(1)}(\mathbf{x}, t) + \sum_j ([D^{(2)}(\mathbf{x}, t)]^{1/2})_{ij} \Gamma_j(t) . \quad (1.15)$$

1.3 Stochastic modelling of experimental data

Since its introduction in [Siegert et al 1998], the above presented formalism of estimating drift and diffusion coefficients directly from a set of experimental data has been applied to a wide range of more or less physical systems. In the review [Friedrich et al 2008], these applications are basically divided into the two categories of complexity in time and complexity in scale. In the first case, one faces the analysis of temporal disorder and for the second case one considers disorder in scale that is often linked to a fractal scaling behaviour. Such a scaling behaviour is found in turbulent flows as well as in financial data (see e.g. [Renner 2002]). A further example for the investigation of complexity in scale is the analysis of surface roughnesses as described in [Wächter et al 2004]. In this thesis, I do not further address the issue of complexity in scale in more detail but concentrate on the issue of complexity in time. The study of complexity in time has covered so far an even wider range of applications including the analysis of human movement [Friedrich et al 2000, Frank et al 2006, Van Mourik et al 2006], biological data as heart-rate fluctuations [Kuusela et al 2003, Ghasemi et al 2005, Tabar et al 2006] and brain dynamics [Prusseit & Lehnertz 2007], the modelling of electronic circuits [Friedrich et al 2000, Stemler et al 2007], diverse price dynamics [Ghasemi et al 2007, Farahpour et al 2007] as well as pattern formation in the broadest sense like ocean variability and sea surface winds [Sura & Gille 2003, Lind et al 2005], the dynamics of traffic flow [Kriso et al 2002] and the segregation behaviour of granular material [Kern et al 2005].

The data that is considered for the analysis most often corresponds to a simple discrete time series of the measured physical quantity. In some cases, however, the reconstruction procedure is applied to a modified time series as defined e.g. by the returns or the increments of the sampled data to overcome the problem of non-stationarity of the respective data set (cf. [Tabar et al 2006, Ghasemi et al 2007]). The stationarity of the data that is analyzed is a basic requirement of the modelling approach.

The general procedure of the above mentioned and similar applications consists of three steps. The first step is to verify the properties that are necessary, beside the stationarity of the data set, for the data to be modelled as a Langevin process. The second step is the actual estimation of the drift and diffusion coefficients. Eventually, as a last step a kind of self-consistency check is performed, comparing the statistical properties of the data that is simulated on the basis of the reconstructed

effective Langevin equation with the original empirical data to verify the estimated coefficients. Not all of the listed applications actually follow this scheme but it can be seen as a general recommendation.

For the first step, the Markov property of the considered process as well as the condition $D^{(4)}(\mathbf{x}, t) = 0$ have to be verified. The most straightforward way to show the Markov property for a certain set of experimental data is to directly evaluate its definition (1.4). Due to the finite amount of data points, the multiple conditioning on the left side is usually reduced to only two conditions. The corresponding double conditioned pdf is compared to the single conditioned pdf on the right side either by visual comparison or by means of a statistical test (cf. [Renner 2002, Wächter et al 2004]). Another method to verify Markov properties is to show the validity of the Chapman-Kolmogorov equation

$$p(\mathbf{x}_j, t_j | \mathbf{x}_i) = \int d\mathbf{x}_k p(\mathbf{x}_j, t_j | \mathbf{x}_k, t_k) p(\mathbf{x}_k, t_k | \mathbf{x}_i, t_i) \quad (1.16)$$

which is a direct consequence of (1.4). In [Friedrich et al 1998], the validity of this equation is utilized to define a length scale l_{mar} , in later publications referred to as Markov length scale, that defines the smallest time scale with respect to that the process can be shown to be a Markov process. As presented e.g. in [Tabar et al 2007], this length scale may also be utilized as a critical quantity to characterize a certain process and to identify characteristic changes in its behaviour. A third way to investigate the Markov property of a set of experimental data is to firstly derive the underlying effective Langevin equation according to (1.10) and then reconstruct the noise process $\mathbf{\Gamma}(t)$ and inspect its properties. This indirect approach is especially a good alternative for short data sets since no conditioning is necessary (see e.g. [Kern et al 2005] for an application). If a process does not fulfil the Markov property, it can be turned into a Markov process by introducing new variables and hereby increasing the dimensionality of the stochastic state variable (cf. [Risken 1989]).

An essential point for the estimation of the Kramers-Moyal coefficients is the performance of the limit $\lim_{\tau \rightarrow 0} M^{(n)}(\mathbf{x}, t, \tau)/\tau$ in (1.10). A common procedure is to evaluate the conditional moments for the smallest available finite time step τ and then extrapolate them to $\tau = 0$ – as explicitly stated e.g. in [Friedrich et al 2000] and [Prusseit & Lehnertz 2007]. For the most other applications that are listed above no comment concerning this point is found. But I assume that, in a majority of cases, a linear extrapolation is performed likewise. In [Sura & Barsugli 2002], however, it is argued that such an extrapolation may lead to systematic errors for the drift and diffusion coefficients. To avoid these deviations, the coefficients should be evaluated on the basis of an Itô-Taylor expansion for the conditional moments. In [Kleinhans et al 2005], this argument is taken up, and an iterative procedure for the estimation of drift and diffusion coefficients is proposed by adjusting the estimates iteratively to the conditional pdf of the empirical data by the use of the corresponding Kullback-Leibler distance between the empirical and the reconstructed pdfs. In Chapter 2 of this thesis, I resume this discussion and propose an alternative, more direct, approach to optimize drift and diffusion estimates.

A general question is to what extent the theoretical model of a Langevin process actually describes the process under investigation. What I want to point out here is that corresponding deviations and restrictions do not mean that the presented reconstruction scheme is completely useless for the characterization of the considered process. In [Siefert & Peinke], it is e.g. shown that even if the stochastic variable $\mathbf{\Gamma}(t)$ in a stochastic differential equation as defined by (1.2) or (1.15), respectively, refers to neither Gaussian distributed nor δ -correlated noise, the function $D^{(1)}(\mathbf{x}, t)$ describing the deterministic part of the dynamics can be reconstructed correctly. The corresponding stochastic differential equation should not be called Langevin

equation in this case, and it is not clear how to define a diffusion in this context. Nevertheless, this result is promising. The self-consistency test, proposed as verification of the estimated coefficients, however loses in this case its significance. In the same way, the validation of the Markov properties is no longer a pre-condition for the modelling procedure. The Langevin approach can be extended even further by concentrating not on the entire deterministic dynamics but reduce the analysis to the fixed points of the dynamics. This issue is discussed in Chapter 3 of this thesis.

A point that I have not discussed so far is the impact of external measurement noise on the reconstruction procedure. In contrast to dynamical noise that directly interferes with the deterministic dynamics as defined in the Langevin equation, measurement noise is so to say subsequently superimposed on the dynamics. In [Renner et al 2001], [Siefert et al 2003] and [Böttcher et al 2006], it is shown how the presence of measurement noise is expressed in the conditional moments of the disturbed data and how the intrinsic drift and diffusion coefficients can in principle be reconstructed utilizing this information. At this, a filtering of the data prior to the analysis is not necessary but the external noise is extracted within the procedure of reconstruction. I pick up this issue in Chapter 2 of this thesis with a further development of the procedure proposed in [Böttcher et al 2006].

In [Kleinhans et al 2007], it is shown that measurement noise may spoil the Markov properties of an underlying Markov process. With this result in mind, a failed Markov test needn't mean that the process under investigation cannot be described as a Langevin process and the presented method of reconstruction cannot be applied in the strict sense. Therefore, the verification of Markov properties as first step of the analysis must be considered also here carefully.

1.4 Extension to Langevin-like processes

As already indicated, the Langevin approach may give a framework for more general or in some way modified stochastic processes. In this section, I want to give three specific examples for such Langevin-like processes and briefly outline how their dynamics can be reconstructed.

The first example is the modelling of nonlinear Lévy processes as described in [Siegert & Friedrich 2001]. Writing (1.2) in the differential form

$$d\boldsymbol{\xi}(t) = \mathbf{g}(\boldsymbol{\xi}, t)dt + \mathbf{h}(\boldsymbol{\xi}, t)d\mathbf{W}(t) \quad (1.17)$$

and replacing the Langevin force by the more general Lévy noise, the differential equation

$$d\boldsymbol{\xi}(t) = \mathbf{g}(\boldsymbol{\xi}, t)dt + \mathbf{h}(\boldsymbol{\xi}, t)d\mathbf{L}_\alpha^{(\gamma, \beta, \mu)} \quad (1.18)$$

is obtained where $d\mathbf{L}_\alpha^{(\gamma, \beta, \mu)}$ stands for an infinitesimal α -stable Lévy motion with

$$(d\mathbf{L}_\alpha^{(\gamma, \beta, \mu)})_i = (\mathbf{f}_\alpha^{\gamma, \beta, \mu})_i(t)dt. \quad (1.19)$$

The components $(\mathbf{f}_\alpha^{\gamma, \beta, \mu})_i$ of the Lévy noise are assumed to be stochastically independent and characterized by the Lévy stability parameter α , the scale parameter γ , the skewness β and the center μ . The Langevin equation (1.17) is a special case of (1.18) for $\alpha = 2$, $\gamma = 0.5$ and $\beta = \mu = 0$. For simplicity, in [Siegert & Friedrich 2001] a model is assumed in which the components $(\mathbf{f}_\alpha^{\gamma, \beta, \mu})_i$ are no longer necessarily stochastically independent and $\mathbf{h}(\boldsymbol{\xi}, t)$ is replaced by a function $\tilde{\mathbf{h}}(\boldsymbol{\xi}, t)$ with $\tilde{h}_{ij}(\boldsymbol{\xi}, t) = \tilde{h}_{ii}(\boldsymbol{\xi}, t)\delta_{ij}$ and $\tilde{h}_{ii} > 0$, i.e. neglecting cross correlations between the single noise components. Considering furthermore a Lévy noise with $\gamma = 1$, $\beta = \mu = 0$ and $\alpha \in (1, 2]$, $\mathbf{g}(\mathbf{x})$ can be reconstructed similarly as for a Langevin

process by calculating the first conditional moments and performing the limit $\lim_{\tau \rightarrow 0}(\dots)/\tau$. To reconstruct the stochastic part of the process, firstly α has to be estimated and then the components $\hat{h}_{ii}(\mathbf{x})$ can be derived in a straightforward way, having in mind that the noise force of a general Lévy process scales with $\tau^{1/\alpha}$. Thus, the presented method of reconstructing drift and diffusion is extended to processes with more general noise sources without the need for a totally new approach.

The second example I want to focus on is the class of stochastic delay processes, i.e. processes that are described by a time-delayed state variable, as discussed e.g. in [Frank et al 2003, Frank et al 2004]. Dynamical processes involving time delays are particularly found in biological systems that exhibit time-delayed control mechanisms arising due to the transport of matter, energy and information with finite propagation velocities. A general description for this class of systems is given by the stochastic differential equation

$$\dot{\xi}(t) = h(\xi, \xi_\tau, t) + g(\xi, \xi_\tau)\Gamma(t) \quad (1.20)$$

where ξ_τ denotes the time-delayed state variable $\xi_\tau(t) \equiv \xi(t - \tau)$. To simplify matters, ξ and ξ_τ are assumed to be one-dimensional variables. $\Gamma(t)$ is defined as Langevin force corresponding to Gaussian distributed and δ -correlated white noise. Due to considering delayed and non-delayed state variables at the same time, (1.20) does not describe a Markov process. To treat a stochastic delay process, however, in the same way as a Langevin process and apply the presented reconstruction scheme, it can be formulated as a multivariate process without delay. This approach is called method of steps (cf. [Frank et al 2003]). The conclusion to express univariate non-Markovian processes in terms of multivariate Markov processes is in line with the already mentioned fact that non-Markovian processes can be expressed in terms of Markov processes by introducing new state variables.

As last example for a Langevin-like process, I want to mention the effective potential model for the price dynamics of a stock that is introduced and discussed in [Alfi et al 2007]. Here, it is proposed that price dynamics can be described by the model

$$P(t+1) = P(t) - b(t) \frac{d}{dP(t)} \Phi(P(t) - P_M(t)) + \omega(t), \quad (1.21)$$

i.e. in terms of a stochastic process that includes a deterministic force with respect to the moving average $P_M(t)$ of the considered prize variable. $\omega(t)$ is assumed to be a random noise with unitary variance and zero mean similar to a Langevin force but without requiring the Gaussian distribution explicitly. The potential Φ together with the pre-factor $b(t)$ describes the interaction between the prize and its moving average. In the simple assumption of a linear force, Φ results to be a quadratic function according to

$$\Phi(P(t) - P_M(t)) = (P(t) - P_M(t))^2. \quad (1.22)$$

Different regimes of the prize dynamics of a certain stock are characterized in terms of an attractive or repulsive behaviour depending on the sign of $b(t)$ in each case. The deterministic force in (1.21) is reconstructed in a similar way as the drift coefficient for the Langevin approach though clearly not describing the deterministic part of an actual Langevin process. The stochastic part of the equation is not further considered. This example again illustrates that one does not have to require Markov properties to reconstruct the deterministic dynamics of a stochastic process utilizing the Langevin approach or at least a procedure that is adjusted to this approach.

In conclusion, the three presented examples illustrate that the approach of the stochastic modelling, I refer to in this thesis, may be applied to a wide range of

different systems, and that the respective description as a Langevin or Langevin-like process may be understood as a practical framework rather than as an ultimate restriction. In the following, I utilize this framework to analyze and evaluate the data of two specific experimental applications, dealing at the same time with the general aspect of the applicability of the introduced concepts.

1.5 Outline of this thesis

This thesis deals with the modelling of complex systems in terms of stochastic processes that are described by effective Langevin equations. In particular, it is focused on the aspect of application and the respective procedure of reconstruction is illustrated by means of two different examples. Hereby, I concentrate not only on utilizing the technical procedure of applying the theoretical concept to a certain set of experimental data but also investigate in a more general framework an evaluation of applicability. At this, I discuss in detail the discrepancies between the observed dynamical behaviour for a specific application and the ideal dynamics that is defined by the corresponding well-defined Langevin process assumed as theoretical basis.

The thesis is partly written in a cumulative way, including three publications that have been in a slightly modified version already published in or, respectively, submitted to an international scientific journal. In detail, it is organized as follows.

Chapter 1 has given a general outline of the used method of reconstruction and an overview of a large variety of different applications. In Chapter 2, I deal with the question how drift and diffusions estimates, that may define an effective Langevin equation, can be derived for processes that are sampled with only a low resolution and/or affected by additional measurement noise. In particular, it is shown how these estimates differ from the intrinsic drift and diffusion functions that determine the actual dynamics of the considered process. Special attention is, in this connection, paid to the reconstruction of fixed points as characteristic states for a certain process. This discussion is resumed in the context of the utilization of a stochastic fixed-point analysis in Chapter 3. The results of Chapter 2 form the theoretical or rather methodic main part of this thesis whereas Chapter 3 and 4 deal with specific applications. In Chapter 3, I introduce on the basis of a theoretical Langevin model a phenomenological approach to the modelling of a wind turbine's power performance. Particular attention is paid to the description of turbulent structures in wind speed and respective power output time series, forming a self-contained section, as well as to the definition of a stochastic fixed-point analysis as characterization of a certain process. As a second application, Chapter 4 deals with the stochastic modelling of human postural control. With this example, I especially show how the presented reconstruction scheme can be adapted to quite short data sets. Chapter 5 completes this thesis with a summary of the main results and respective final remarks.

A bibliography and an optional set of appendices are given after each chapter, corresponding to the cumulative organization of this thesis.

Bibliography

- [Alfi et al 2007] V. Alfi, A. De Martino, L. Pietronero, and A. Tedeschi: Detecting the traders' strategies in minority-majority games and real stock-prices. In: *Physica A* 382, 1–8 (2007).
- [Böttcher et al 2006] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase: Reconstruction of Complex Dynamical Systems Affected by Strong Measurement Noise. In: *Physical Review Letters* 97, 090603 (2006).
- [Farahpour et al 2007] F. Farahpour, Z. Eskandari, A. Bahraminasab, G. R. Jafari, F. Ghasemi, M. Sahimi, and M. R. R. Tabar: A Langevin Equation for the Rates of Currency Exchange Based on the Markov Analysis. In: *Physica A* 385:2, 601–608 (2007).
- [Frank et al 2003] T. D. Frank, P. J. Beek, and R. Friedrich: Fokker-Planck perspective on stochastic delay systems: Exact solutions and data analysis of biological systems. In: *Physical Review E* 68, 021912 (2003).
- [Frank et al 2004] T. D. Frank, P. J. Beek, and R. Friedrich: Identifying noise sources of time-delayed feedback systems. In: *Physics Letters A* 328:2–3, 219–224 (2004).
- [Frank et al 2006] T. D. Frank, R. Friedrich, and P. J. Beek: Stochastic order parameter equation of isometric force production revealed by drift-diffusion estimates. In: *Physical Review E* 74, 051905 (2006).
- [Friedrich et al 1998] R. Friedrich, J. Zeller, and J. Peinke: A note on three-point statistics of velocity increments in turbulence. In: *Europhysics Letters* 41(2), 153–158 (1998).
- [Friedrich et al 2000] R. Friedrich, S. Siegert, J. Peinke, St. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister: Extracting model equations from experimental data. In: *Physics Letters A* 271, 217–222 (2000).
- [Friedrich et al 2008] R. Friedrich, J. Peinke, and M. R. R. Tabar: Importance of fluctuations: Complexity in the view of stochastic processes. In: *Encyclopedia of Complexity and System Science*, ed B. Meyers; Springer, Berlin (2008).
- [Ghasemi et al 2005] F. Ghasemi, J. Peinke, M. Sahimi, and M.R. Rahimi Tabar: Regeneration of stochastic processes: an inverse method. In: *The European Physical Journal B* 47, 411–415 (2005).
- [Ghasemi et al 2007] F. Ghasemi, M. Sahimi, J. Peinke, R. Friedrich, G. R. Jafari, and M. R. R. Tabar: Analysis of Nonstationary Stochastic Processes with Application to the Fluctuations in the Oil Price. In: *Physical Review E*, 060102(R) (2007).
- [Haken 1983] H. Haken: *Synergetics*. Springer, Berlin (1983).
- [Kantz & Schreiber 1997] H. Kantz, T. Schreiber: *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, England (1997).
- [Kern et al 2005] M. Kern, O. Buser, J. Peinke, M. Siefert, and L. Vulliet: Stochastic analysis of single particle segregational dynamics. In: *Physics Letters A* 336:4–5, 428–433 (2005).
- [Kleinhans et al 2005] D. Kleinhans, R. Friedrich, A. Nawroth, and J. Peinke: An iterative procedure for the estimation of drift and diffusion coefficients of Langevin processes. In: *Phys. Lett. A* 346, 42–46 (2005).

- [Kleinhans et al 2007] D. Kleinhans, R. Friedrich, M. Wächter, and J. Peinke: Markov properties in presence of measurement noise. In: *Physical Review E* 76, 041109 (2007).
- [Kolmogorov 1931] A. N. Kolmogorov: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. In: *Mathematische Annalen* 104, 415–458 (1931).
- [Kriso et al 2002] S. Kriso, R. Friedrich, J. Peinke, and P. Wagner: Reconstruction of dynamical equations for traffic flow. In: *Physics Letters A* 299:2-3, 287–291 (2002).
- [Kuusela et al 2003] T. Kuusela, T. Shephard, and J. Hietarinta: Stochastic model for heart-rate fluctuations. In: *Physical Review E* 67, 061904 (2003).
- [Lind et al 2005] P. G. Lind, A. Mora, J. A. C. Gallas, and M. Haase: Reducing stochasticity in the North Atlantic index with coupled Langevin equations. In: *Physical Review E*, 056706 (2005).
- [Prusseit & Lehnertz 2007] J. Prusseit and K. Lehnertz: Stochastic Qualifiers of Epileptic Brain Dynamics. In: *Physical Review Letters* 98, 138103 (2007).
- [Renner et al 2001] Ch. Renner, J. Peinke, and R. Friedrich: Evidence of Markov properties of high frequency exchange rate data. In: *Physica A* 298, 499–520 (2001).
- [Renner 2002] Ch. Renner: *Markowanalysen stochastisch fluktuierender Zeitserien*. PhD thesis, University of Oldenburg (2002).
- [Risken 1989] H. Risken: *The Fokker-Planck Equation*. Springer, New York (1989).
- [Siefert et al 2003] M. Siefert, A. Kittel, R. Friedrich, and J. Peinke: On a quantitative method to analyze dynamical and measurement noise. In: *Europhysics Letters* 61:4, 466–472 (2003).
- [Siefert & Peinke] M. Siefert and J. Peinke: Reconstruction of the deterministic dynamics of stochastic systems. In: *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering* 14:6, 2005–2010 (2004).
- [Siegert et al 1998] S. Siegert, R. Friedrich, and J. Peinke: Analysis of data sets of stochastic systems. In: *Phys. Lett. A* 243, 275–280 (1998).
- [Siegert & Friedrich 2001] S. Siegert and R. Friedrich: Modeling of nonlinear Lévy processes by data analysis. In: *Physical Review E* 64, 041107 (2001).
- [Stemler et al 2007] T. Stemler, J. P. Werner, H. Benner, and W. Just: Stochastic modelling of experimental chaotic time series. In: *Physical Review Letters* 98, 044102 (2007).
- [Sura & Barsugli 2002] P. Sura and J. Barsugli: A note on estimating drift and diffusion parameters from timeseries. In: *Physics Letters A* 305, 304–311 (2002).
- [Sura & Gille 2003] P. Sura and S. T. Gille: Interpreting wind-driven Southern Ocean variability in a stochastic framework. In: *Journal of Marine Research* 61:3, 313–334 (2003).
- [Tabar et al 2006] M. R. R. Tabar, F. Ghasemi, J. Peinke, R. Friedrich, K. Kavianian, F. Taghavi, S. Sadeghi, G. Bijani, and M. Sahimi: New computational approaches to analysis of interbeat intervals in human subjects. In: *Computing in Science and Engineering*, 86–97 (2006).

- [Tabar et al 2007] M. R. R. Tabar, M. Sahimi, K. Kaviani, M. Allamehzadeh, J. Peinke, M. Makhtari, M. Vesaghi, M. D. Niry, F. Ghasemi, A. Bahraminasab, S. Tabatabai, and F. Fayazbakhsh: Dynamics of the Markov Time Scale of Seismic Activity May Provide a Short-Term Alert for Earthquakes. In: *Modelling Critical and Catastrophic Phenomena in Geoscience: A Statistical Physics Approach*. p. 281-301. Springer, Berlin-Heidelberg (2007).
- [Van Mourik et al 2006] A. M. van Mourik, A. Daffertshofer, and P. J. Beek: Deterministic and stochastic features of rhythmic human movement. In: *Biological Cybernetics* 94:3, 233–244 (2006).
- [Wächter et al 2004] M. Wächter, F. Riess, T. Schimmel, U. Wendt, and J. Peinke: Stochastic analysis of different rough surfaces. In: *The European Physical Journal B* 41, 259–277 (2004).

Chapter 2

On the definition and handling of different drift and diffusion estimates ¹

A previously devised approach for the reconstruction of Langevin processes from given data is revised with respect to disturbances stemming from finite sampling rates and the presence of external measurement noise. For these two cases and a combination of both three different estimates for the drift and diffusion functions are introduced, and an optimization procedure is presented that allows the reconstruction of the intrinsic functions from these estimates. Special attention is paid to the reconstruction of deterministic fixed points defining the characteristic behaviour of a process, and its robustness against the considered disturbing effects.

2.1 Introduction

Complex dynamical systems appear in various scientific disciplines, and a suitable mathematical description is often of central interest. Revealing the process dynamics is realized by reconstructing the corresponding dynamical equations and finding a model for the dynamics of the studied systems. Prior to the reconstruction, the analysis is usually constrained to a special class of processes. A fundamental distinctive feature is e.g. whether one applies a deterministic or a stochastic framework. Observed complex fluctuations can thus be traced back either to a chaotic behaviour or the effect of dynamical noise. A more flexible approach is to combine both features and analyze a system in terms of general stochastic processes.

Here, we consider the class of Langevin processes, i.e. stochastic processes that are assumed to be Markovian and described in terms of drift and diffusion. At this, we only consider autonomous processes, i.e. the corresponding governing coefficients are explicitly not time-dependent. A one-dimensional Langevin process is thus given by the stochastic first-order differential equation

$$\dot{x}(t) = D^{(1)}(x) + \sqrt{D^{(2)}(x)} \Gamma(t) \quad (2.1)$$

(also called Langevin equation). Even though we restrict our discussion here to the one-dimensional case, it is straightforward to extend the argumentation to

¹Published as J. GOTTSCHALL and J. PEINKE: On the definition and handling of different drift and diffusion estimates. In: *New Journal of Physics* 10, 083034 (2008).

higher dimensional processes. The term $D^{(1)}(x)$ is named drift coefficient and reflects the deterministic part of the dynamics. The stochastic part is given by the Langevin force $\Gamma(t)$, representing Gaussian white noise with $\langle \Gamma(t) \rangle = 0$ and $\langle \Gamma(t_1)\Gamma(t_2) \rangle = 2\delta(t_1-t_2)$ (following the convention in [Risken 1989]), and the square root of the diffusion coefficient $D^{(2)}(x)$, fixing the amplitude of the stochastic fluctuations. Throughout the paper, we apply Itô's interpretation of stochastic integrals.² Equation (2.1) is directly connected to the Fokker-Planck equation

$$\frac{\partial}{\partial t}p(x, t) = \left[-\frac{\partial}{\partial x}D^{(1)}(x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}D^{(2)}(x) \right] p(x, t), \quad (2.2)$$

describing the process in probability space.

In recent years, a method has been introduced to reconstruct the coefficients $D^{(n)}(x)$ ($n = 1, 2$) directly from a measured time series [Siegert et al 1998, Friedrich & Peinke 1997, Ryskin 1997, Friedrich et al 2000]. As initially proposed by Siegert et al., drift and diffusion coefficient can be derived according to

$$D^{(n)}(x) = \frac{1}{n!} \lim_{\tau \rightarrow 0} \frac{1}{\tau} M^{(n)}(x, \tau) \quad (2.3)$$

with the first ($n = 1$) and second ($n = 2$) conditional moment, respectively, defined by

$$M^{(n)}(x, \tau) := \langle [x(t+\tau) - x(t)]^n \rangle |_{x(t)=x}. \quad (2.4)$$

For an ideal time series, i.e. a time series $x(t)$ that is generated by a process given by (2.1) and sampled over a sufficiently long time period and with a sufficiently high resolution, the process dynamics can be perfectly reconstructed by (2.3) and (2.4). For real data sets, however, the method of reconstruction must be revised with respect to several kinds of aspects – three of them are discussed in the following.

An observed time series $x(t)$ is defined by a finite sampling rate. Thus, the continuous process given by (2.1) is reduced to a series of discrete points separated by the time increment τ_0 corresponding to the inverse sampling rate, and information from smaller time scales is ignored (see figures 2.1(a)–(b)). Equation (2.1) is replaced by the discrete equation

$$x(t+\tau) = x(t) + \tau D^{(1)}(x) + \sqrt{\tau D^{(2)}(x)} \Gamma'(t), \quad (2.5)$$

where the $\Gamma'(t)$ correspond to independent Gaussian-distributed random numbers with zero mean and variance 2. Applying a conditional ensemble average and using the properties of $\Gamma'(t)$ specified above, we obtain as finite- τ approximations

$$M^{(1)}(x, \tau) \approx \tau D^{(1)}(x), \quad (2.6)$$

$$M^{(2)}(x, \tau) \approx 2\tau D^{(2)}(x) + \tau^2 (D^{(1)}(x))^2. \quad (2.7)$$

That is, $M^{(2)}(x, \tau)$ is modified by a correction term that is quadratic in $D^{(1)}(x)$ and τ and that vanishes in the limit $\tau \rightarrow 0$. In [Friedrich et al 2002] it is shown that additional terms for $M^{(2)}(x, \tau)$ as well as for $M^{(1)}(x, \tau)$ are obtained when deriving the τ -dependence of the conditional moments directly from the Fokker-Planck equation. For the approximations (2.6) and (2.7), however, a simplified transition probability is assumed (cf. [Ragwitz & Kantz 2001]). Ignoring any $\mathcal{O}(\tau^2)$ -corrections and using the simple relation $D^{(n)}(x) = M^{(n)}(x, \tau_0)/(n!\tau_0)$ for the analysis of non-ideal data,

²It is just as well possible to reformulate the approach using other interpretations of stochastic integrals, for instance Stratonovich's definition (cf. [Risken 1989]). Applying this definition, the drift coefficient is altered by an additional term, the spurious drift, and the reconstruction scheme is adapted correspondingly. But in total we do not expect to obtain fundamentally new insights.

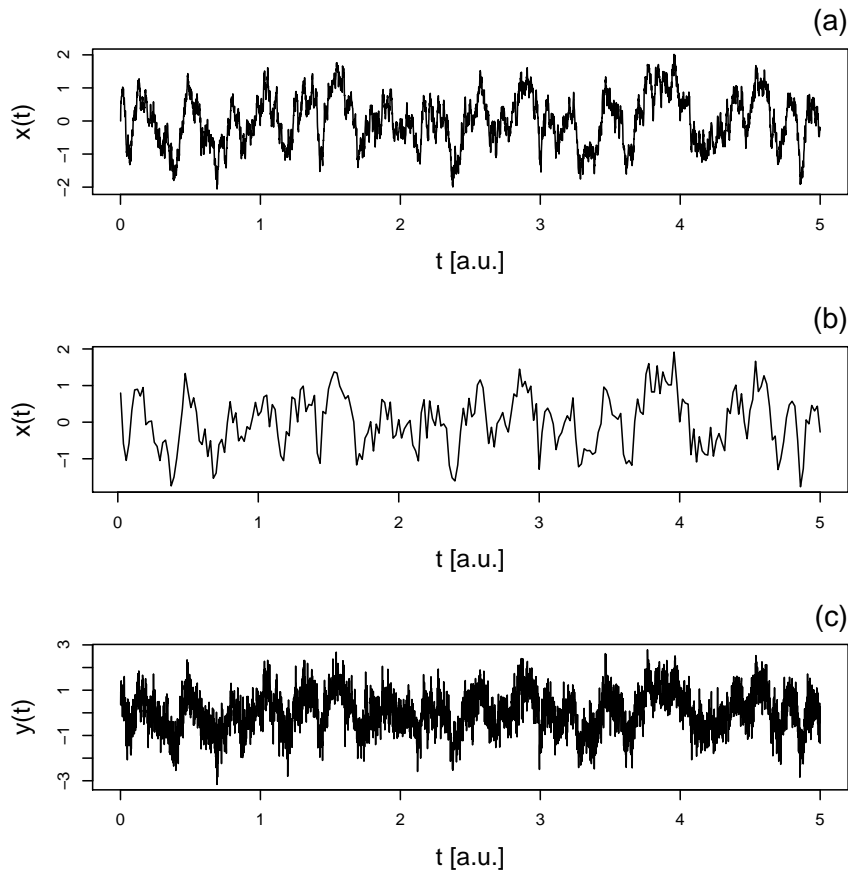


Figure 2.1: Time series of an exemplary stochastic process (Ornstein-Uhlenbeck process as defined below) – (a) sampled with $\tau_0 = 10^{-3}$, (b) $\tau_0 = 0.2$, and (c) superimposed by measurement noise ($\sigma = 0.5$, $\tau_0 = 10^{-3}$).

the empirical estimates may significantly differ from the intrinsic functions $D^{(n)}(x)$ ($n = 1, 2$) (cf. [Sura & Barsugli 2002]).

Similar deviations are observed when the time series $x(t)$ is superimposed by external measurement noise – see figure 2.1(c). The respective correction terms can be derived likewise, as presented in [Böttcher et al 2006]. It is important to note that in both cases, for a not sufficiently high resolution as well as in the presence of external noise, the actual dynamics of the considered Langevin process is merely hidden but not destroyed. The challenge is to define corresponding estimates that are adjusted to the individual disturbances and to interpret the respective results in an appropriate way, or to introduce an appropriate optimization scheme that allows to reconstruct the intrinsic functions from the estimates.

The situation is a bit different, and this is the third aspect we address, if the requirement of the Markov property for the process given by $x(t)$ is not strictly fulfilled. Following [Risken 1989], this is the case if the stochastic force $\Gamma(t)$ in (2.1) does not correspond to Gaussian distributed and especially δ -correlated white noise. For such non-Markovian processes, (2.3) and (2.4) can still be used to reconstruct the deterministic part of the process dynamics. Experimental evidence has been presented in [Siefert & Peinke 2004] where the authors reconstructed the drift coefficient $D^{(1)}(x)$ properly according to (2.3) for noise sources that are not Langevin

forces and especially not δ -correlated.

Allowing explicitly the non-Markovianity and non-Gaussianity of the noise source, the presented reconstruction scheme can be extended to a broader class of processes while its relevance is reduced to the deterministic part of the dynamics. A further extension, associated with a further reduction of significance, is achieved by considering not the entire deterministic dynamics of the process but just focusing on its characteristic or say global behaviour. For instance, it might be interesting to reconstruct only the deterministic fixed points x_{FP} of an observed process. These are defined by

$$D^{(1)}(x_{\text{FP}}) \equiv 0. \quad (2.8)$$

The derivative of $D^{(1)}(x)$ in the neighbourhood of x_{FP} determines the type of local stability. For a stable fixed point $(D^{(1)}(x))'|_{x=x_{\text{FP}}}$ is negative, for an unstable fixed point positive. A respective application is given in [Anahua et al 2008]. Here, the characteristic power curve of a wind turbine, i.e. the curve of steady states of power output for certain wind speeds, is determined by reconstructing the deterministic fixed points of the process binned according to the speed. We could show in [Gottschall & Peinke 2008] that the fixed points are more appropriate to characterize the power performance of the wind turbine system than maxima or mean values which are commonly used in the standard procedures. The question however arises how robust fixed points are against spoiling like the presence of external noise and finite sampling rates. An answer will be given at the end of this chapter. Further applications are given in [Kriso et al 2002] and [Kern et al 2005] where the authors analyzed traffic flows and the segregation dynamics in avalanches, respectively, with similar approaches.

This chapter is arranged as follows. Subsequent to this introduction, we give an overview of the finite time and measurement noise corrections introduced in [Ragwitz & Kantz 2001] and [Böttcher et al 2006], respectively. In section 2.4 we compare both effects with each other, define distinctive features and introduce a scheme to combine them. Thereby, we develop three different estimates that are utilized for an optimized reconstruction of the intrinsic functions in section 2.5 and discussed with respect to their significance, especially with regard to a proper reconstruction of fixed points, in section 2.6. We conclude with section 2.7 which summarizes the main results.

2.2 Definition of drift and diffusion estimates adapted to finite sampling rates

As annotated in [Ragwitz & Kantz 2001], the estimation of drift and diffusion coefficients for experimental data generally suffers from finite sampling rates, i.e. a finite time increment τ_0 . In [Friedrich et al 2002] respective corrections are introduced by deriving exact expressions for the conditional moments up to a specified order of τ from the Fokker-Planck equation. In the following, we summarize these considerations before interpreting and applying the results.

The conditional moments $M^{(n)}(x, \tau)$ ($n = 1, 2$), as defined in (2.4), are formulated according to

$$\langle [x(t + \tau) - x(t)]^n \rangle_{|x(t)=x} = \int dx (x - x')^n p(x, t + \tau | x', t) \quad (2.9)$$

in terms of the conditional probability density function (pdf) $p(x, t + \tau | x', t)$. This conditional pdf is the solution of the Fokker-Planck equation

$$\frac{\partial}{\partial t} p(x, t + \tau | x', t) = \hat{L} p(x, t + \tau | x', t) \quad (2.10)$$

with the Fokker-Planck operator

$$\hat{L} = -\frac{\partial}{\partial x}D^{(1)} + \frac{\partial^2}{\partial x^2}D^{(2)}. \quad (2.11)$$

The formal solution of (2.10) reads

$$p(x, t + \tau | x', t) = \exp[\hat{L}\tau]\delta(x - x'). \quad (2.12)$$

Expanding it in a stochastic Itô-Taylor series yields

$$p(x, t + \tau | x', t) = \sum_{k=0}^{\infty} \frac{(\hat{L}\tau)^k}{k!} \delta(x - x') \quad (2.13)$$

(cf. [Risken 1989]). By inserting (2.13) into (2.9), we obtain exact expressions for the conditional moments for arbitrary τ or, restricting the expansion (2.13) to some finite order, the respective approximations. In accordance with [Friedrich et al 2002], we find as second-order approximations

$$\begin{aligned} M^{(1)}(x, \tau) &= \tau D^{(1)} + \frac{\tau^2}{2} \left[D^{(1)}(D^{(1)})' + D^{(2)}(D^{(1)})'' \right] \\ &\quad + \mathcal{O}(\tau^3) \end{aligned} \quad (2.14)$$

$$\begin{aligned} M^{(2)}(x, \tau) &= 2\tau D^{(2)} + \tau^2 \left[(D^{(1)})^2 + 2D^{(2)}(D^{(1)})' + D^{(1)} \right. \\ &\quad \left. (D^{(2)})' + D^{(2)}(D^{(2)})'' \right] + \mathcal{O}(\tau^3), \end{aligned} \quad (2.15)$$

where $(\dots)'$ and $(\dots)''$ refer to the first and second derivative of the drift or diffusion coefficient in the parentheses, respectively. In comparison to (2.6) and (2.7), we thus obtain additional terms of order $\mathcal{O}(\tau^2)$.

Following [Kleinhans & Friedrich 2007], we introduce as first estimate, adapted to finite sampling rates,

$$D_{\text{E},\tau}^{(n)}(x, \tau_0) := \frac{M^{(n)}(x, \tau_0)}{n!\tau_0}. \quad (2.16)$$

Inserting the expressions (2.14) and (2.15) with known functions $D^{(1)}(x)$ and $D^{(2)}(x)$ gives a theoretical estimate. An empirical estimate for a specific time series is, on the other hand, obtained with definition (2.4) according to

$$D_{\text{E},\tau}^{(n)}(x, \tau_0) = \frac{1}{n!\tau_0} \langle [x(t + \tau_0) - x(t)]^n \rangle_{|x(t)=x}. \quad (2.17)$$

The connection between both, theoretical and empirical estimate, is directly given by (2.9–2.13).

Turning to the theoretical case, we have analyzed the deviations of the estimates $D_{\text{E},\tau}^{(n)}(x, \tau_0)$ from the original functions for different simple examples. For an Ornstein-Uhlenbeck process (given by $D^{(1)}(x) = -\alpha x$ and $D^{(2)}(x) = \beta$ with $\alpha, \beta = \text{const}$) we obtain as theoretical estimates

$$D_{\text{E},\tau}^{(1)}(x, \tau_0) = -\alpha x \left(1 - \frac{\alpha\tau_0}{2}\right) + \mathcal{O}(\tau_0^2), \quad (2.18)$$

$$D_{\text{E},\tau}^{(2)}(x, \tau_0) = \beta \left(1 - \alpha\tau_0\right) + \frac{\tau_0\alpha^2}{2}x^2 + \mathcal{O}(\tau_0^2). \quad (2.19)$$

The estimate $D_{\text{E},\tau}^{(1)}(x, \tau_0)$ differs from the original function by a decreased slope, the estimate $D_{\text{E},\tau}^{(2)}(x, \tau_0)$ is characterized by an additional quadratic term. For the application, this means that a quadratic diffusion coefficient must not indicate

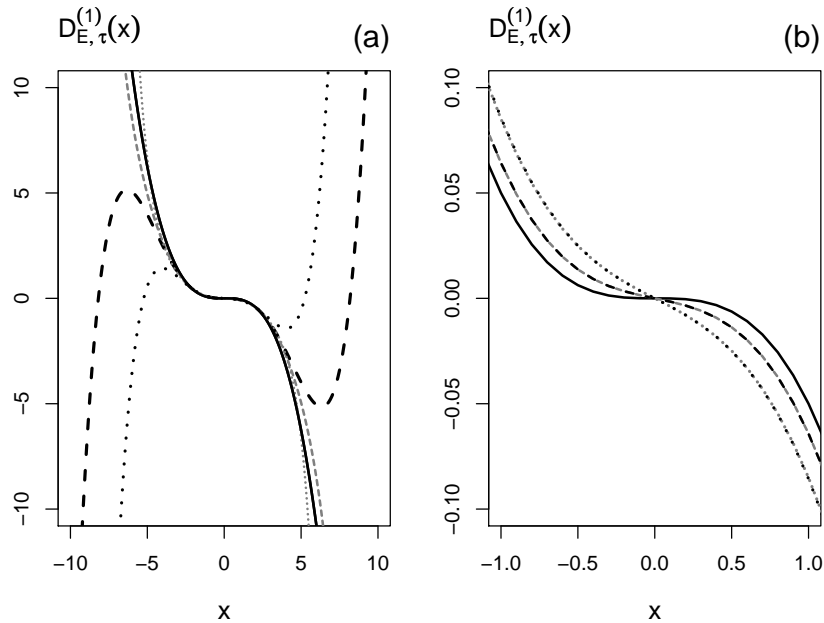


Figure 2.2: The estimate $D_{E,\tau}^{(1)}(x, \tau_0)$ for a process with cubic drift and constant diffusion ($D^{(1)}(x) = -ax^3$ and $D^{(2)} = b$ with $a = 0.05$, $b = 0.5$) for two different values of τ_0 . The dashed ($\tau_0 = 0.2$) and dotted ($\tau_0 = 0.5$) lines indicate second-order estimates due to (2.14), for the grey lines the terms of third order in τ according to (2.44) are additionally considered (again dashed for $\tau_0 = 0.2$ and dotted for $\tau_0 = 0.5$). The original function $D^{(1)}(x)$ is given by the solid black line. Figure (b) is a close-up view of the region around the fixed point at $x = 0$ in (a).

multiplicative noise but can also be caused by a too low sampling rate. While the behaviour of $D^{(2)}(x)$ is thus significantly affected by the finiteness of τ , the behaviour of $D^{(1)}(x)$ changes just quantitatively but not qualitatively. In particular, the fixed point of the process is not affected and can be reconstructed by the estimate correctly.

Next, we consider a process with additive noise ($D^{(2)} = b$) and cubic drift ($D^{(1)}(x) = -ax^3$). Figures 2.2(a-b) show the estimate $D_{E,\tau}^{(1)}(x, \tau_0)$ for two different values of τ_0 (dashed and dotted line, respectively). For both cases, the characteristic behaviour of $D^{(1)}$ is again reproduced correctly by the estimate (2.14) – at least locally. Globally, $D_{E,\tau}^{(1)}(x, \tau_0)$ indicates additional fixed points, to the left and to the right of the original fixed point and with the inverse type of stability. These artefacts sensitively depend on changes of τ_0 , and they disappear when we consider higher-order terms in (2.14). Estimates for the drift coefficient according to (2.16) including additionally the third-order terms of the expansion (2.13) are shown in figures 2.2(a-b) as grey lines. The corresponding third-order corrections to (2.14) are given in the appendix to this chapter. This second example emphasizes that it is of particular importance how many terms we consider for the expansion in (2.13).

To summarize the impact of finite- τ effects, figure 2.3 illustrates exemplarily for an Ornstein-Uhlenbeck process the deviations between the empirical values and the theoretical functions of different orders for $M^{(1)}(x, \tau)$ as well as the deviation between the introduced estimate and the intrinsic function for the drift coefficient. (For the illustration, we restrict our investigations to the analysis of the first con-

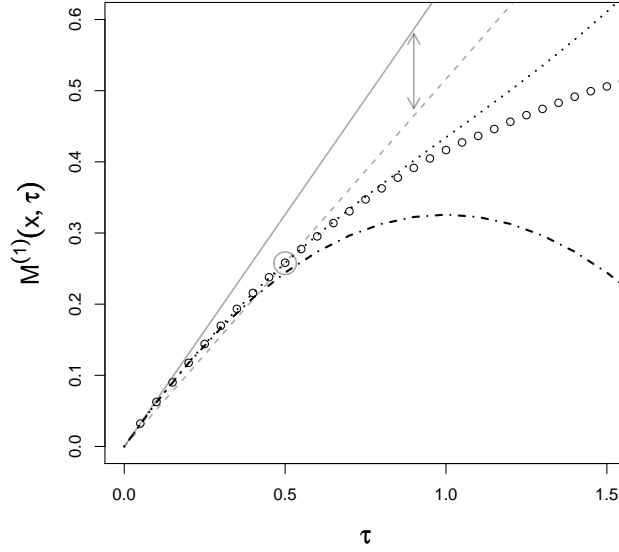


Figure 2.3: Empirical values and theoretical functions for the first conditional moment of an Ornstein-Uhlenbeck process ($\alpha = 1$, $\beta = 1$), and the deviation between the estimate $D_{E,\tau}^{(1)}(x, \tau_0)$ and the intrinsic function $D^{(1)}(x)$. The empirical values for $M^{(1)}(x = -0.7, \tau)$ (according to (2.4)) are given by the symbols, the second- and third-order approximations are denoted by the dot-dashed and dotted lines, resp. The solid line represents the linear approximation $\tau D^{(1)}(x)$, the dashed line the relation $\tau D_{E,\tau}^{(1)}(x, \tau_0 = 0.5)$ with the estimate according to (2.16). The grey circle indicates the value $M^{(1)}(x, \tau_0 = 0.5)$ used for the derivation of the estimate.

ditional moment. An extension to $M^{(2)}(x, \tau)$ is straightforward.) The empirical values for $M^{(1)}(x, \tau)$ (given by the symbols) are characterized by a curvature that is defined by the higher-order terms in τ . Their agreement with the analytical second- and third-order approximations (dotdashed and dotted lines) indicates up to which value of τ the respective approximation is reliable. Furthermore, the illustration shows the deviation between the estimate $D_{E,\tau}^{(1)}(x, \tau_0)$ and the intrinsic function $D^{(1)}(x)$ by comparing the linear relations $\tau D_{E,\tau_0}^{(1)}(x, \tau)$ and $\tau D^{(1)}(x)$. Only in the limit $\tau \rightarrow 0$ they agree, i.e. the linear approximation $M^{(1)}(x, \tau) = \tau D^{(1)}(x)$ is exact, and the larger the value of τ the larger is the deviation between estimate and intrinsic function for $D^{(1)}(x)$.

2.3 Definition of drift and diffusion estimates adapted to the presence of measurement noise

The second disturbance we discuss in this chapter is the presence of external measurement noise. It has already been discussed in [Siefert et al 2003] how adding measurement noise to signals generated from a Langevin process leads to fundamental modifications of the data analysis and the reconstruction of the underlying dynamics. Siefert et al. considered the special case of Gaussian distributed measurement noise defined by $\sigma\zeta(t)$ with $\langle\zeta(t)\rangle = 0$ and $\langle\zeta(t_1)\zeta(t_2)\rangle = \delta(t_1 - t_2)$ and σ as the amplitude of the noise. The presence of external measurement noise means that instead of $x(t)$ the time series $y(t) = x(t) + \sigma\zeta(t)$ is measured, i.e. the intrinsic process is superimposed by the external noise.

Adopting these considerations, Böttcher et al. proposed an advanced approach for the reconstruction of drift and diffusion coefficients based on the calculation of the conditional moments $M_\sigma^{(n)}(y, \tau)$ [Böttcher et al 2006]. Note that $M_\sigma^{(n)}$ is actually the same function as $M^{(n)}$ (see definition (2.4)) but for the time series $y(t)$ instead of $x(t)$. We use this new notation to distinguish the cases with and without external noise and to avoid confusion in this connection.

Again, we summarize the corresponding proceeding, as it was proposed, before we utilize it to define an appropriate estimate. Böttcher et al. pointed out that the coefficients $D^{(n)}(x)$ are implicitly given by the conditional moments of the accessible data $y(t)$ according to ³

$$\begin{aligned} M_\sigma^{(1)}(y, \tau) &= \langle y(t + \tau) - y(t) \rangle |_{y(t)=y=x(t)+\sigma\zeta(t)} \\ &= \tau \int dx D^{(1)}(x) f(x|y) + \int dx (x - y) f(x|y) \\ &= m^{(1)}(y, \tau) + \gamma_1(y) \end{aligned} \quad (2.20)$$

and

$$\begin{aligned} M_\sigma^{(2)}(y, \tau) &= \langle [y(t + \tau) - y(t)]^2 \rangle |_{y(t)=y=x(t)+\sigma\zeta(t)} \\ &= \tau \int dx \left[2(x - y) D^{(1)}(x) + 2D^{(2)}(x) \right] f(x|y) \\ &\quad + \sigma^2 + \int dx (x - y)^2 f(x|y) \\ &= m^{(2)}(y, \tau) + \gamma_2(y), \end{aligned} \quad (2.21)$$

assuming before that the approximation

$$M^{(n)}(x, \tau) \approx \tau n! D^{(n)}(x) + \mathcal{O}(\tau^2) \quad (2.22)$$

for $\tau \ll 1$ holds. The unknown probability density $f(x|y)$ is defined by

$$f(x|y) = \frac{f(y|x)p(x)}{\int dx f(y|x)p(x)}, \quad (2.23)$$

where $f(y|x)$ is the distribution of the measurement noise, i.e. for Gaussian distributed noise

$$f(y|x) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left[-\frac{(y-x)^2}{2\sigma^2}\right], \quad (2.24)$$

and $p(x)$ the distribution of the process $x(t)$, in the stationary case given by

$$p(x) = \frac{\mathcal{N}}{D^{(2)}} \exp\left[\int_{-\infty}^x d\tilde{x} \frac{D^{(1)}(\tilde{x})}{D^{(2)}(\tilde{x})}\right] \quad (2.25)$$

with the normalization constant \mathcal{N} (cf. [Risken 1989]). Equations (2.20) and (2.21) indicate that the presence of measurement noise results in an offset $\gamma_n(y)$ for the conditional moments as function of τ . In order to deal with the divergence of $M_\sigma^{(n)}(y, \tau)/\tau$ resulting from this offset, we suggest to define

$$D_{E,\sigma}^{(n)}(y) := \frac{m^{(n)}(y, \tau)}{n!\tau} \quad (2.26)$$

as second estimate for the drift and diffusion coefficient, respectively, adapted to the presence of external measurement noise. The theoretical estimate is hence

³We introduce the notation $f(\dots)$ here to distinguish the conditioned probability density functions $f(x|y)$ and $f(y|x)$ from the unconditioned probability density $p(x)$.

determined by the first part in (2.20) and (2.21). An empirical estimate is calculated by applying a linear fit to $M_\sigma^{(n)}(y, \tau)$.

For a general process, (2.20) and (2.21) must be evaluated numerically. For an Ornstein-Uhlenbeck process as defined above, however, we can check the effect of the approximation (2.26) analytically. In [Böttcher et al 2006], the terms $\gamma_n(y)$ and $m^{(n)}(y, \tau)$ were derived to

$$\gamma_1(y) = -\frac{\sigma^2}{\lambda^2}y, \quad \gamma_2(y) = \sigma^2 + \frac{\sigma^2 s^2}{\lambda^2} + \frac{\sigma^4}{\lambda^4}y^2 \quad (2.27)$$

and

$$\begin{aligned} m^{(1)}(y, \tau) &= \tau[-\alpha y - \alpha \gamma_1(y)], \\ m^{(2)}(y, \tau) &= 2\tau(\beta - \alpha\{\gamma_2(y) - \sigma^2\} + y\gamma_1(y)) \end{aligned} \quad (2.28)$$

with $\lambda^2 := s^2 + \sigma^2$ and $s^2 = \beta/\alpha$ as the variance of the distribution $p(x)$. It follows

$$D_{E,\sigma}^{(1)}(y) = -\alpha y \left(1 - \frac{\sigma^2}{\lambda^2}\right), \quad (2.29)$$

$$D_{E,\sigma}^{(2)}(y) = \beta \left(1 - \frac{\sigma^2}{\lambda^2}\right) + \alpha \left(\frac{\sigma^2}{\lambda^2} - \frac{\sigma^4}{\lambda^4}\right) y^2. \quad (2.30)$$

That is, in the presence of measurement noise, similar to the case of a finite sampling rate, the estimate for $D^{(1)}(x)$ differs from the original function quantitatively but not qualitatively. The absolute slope α is reduced by a factor depending on σ . The behaviour of $D_{E,\sigma}^{(2)}(y)$, on the other hand, is also qualitatively influenced by the external noise. The constant term β is not only reduced by the same factor depending on σ but additionally superimposed by a quadratic term.

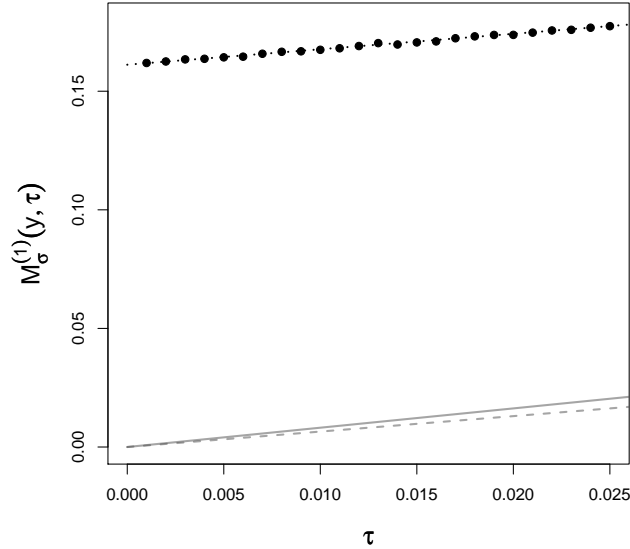


Figure 2.4: Empirical values (symbols) and theoretical function (dotted line) for the first conditional moment $M_\sigma^{(1)}(y = -0.7, \tau)$ of an Ornstein-Uhlenbeck process ($\alpha = 1, \beta = 1$) that is superimposed by measurement noise ($\sigma = 0.5$). The solid line denotes the relation $\tau D^{(1)}(y)$ for the undisturbed case without external noise, the dashed line gives the relation $\tau D_{E,\sigma}^{(1)}(y)$ with the estimate according to (2.26).

For illustration, figure 2.4 compares the theoretical and empirical values for $M_\sigma^{(1)}(y, \tau)$ as well as the estimate $D_{E, \sigma}^{(1)}(y)$ and the undisturbed intrinsic drift function for an Ornstein-Uhlenbeck process – analogous to figure 2.3. The empirical values (given by the symbols) are perfectly reproduced by the theoretical function (dotted line). The deviation between the relation $\tau D_{E, \sigma}^{(1)}(y)$ (dashed line), with the estimate defined as the slope of $M_\sigma^{(1)}(y, \tau)$, and $\tau D^{(1)}(y)$ (solid line) indicate that, though dealing with the offset induced by the presence of external noise, we have not considered the σ -dependence of $m^{(n)}(y, \tau)$ for the estimate $D_{E, \sigma}^{(n)}(y)$ (as given in (2.26)). Note that the application of the estimate $D_{E, \sigma}^{(1)}(y)$ is restricted to small ranges of τ where $M_\sigma^{(1)}(y, \tau)$ vs. τ shows a linear behaviour.

2.4 Drift and diffusion estimates for a combination of low sampling rates and measurement noise

We have analytically shown in the preceding two sections that for an Ornstein-Uhlenbeck process both a finite resolution defined by τ_0 and the presence of external noise may lead to very similar deviations of the introduced estimates from the original functions $D^{(n)}(x)$ ($n = 1, 2$). To illustrate this, figures 2.5(a) and (b) show the empirical values as well as the theoretical functions for $D_{E, \tau}^{(n)}(x, \tau)$ and $D_{E, \sigma}^{(n)}(y)$ for an Ornstein-Uhlenbeck process ($\alpha = 1, \beta = 1$) that is sampled with $\tau_0 = 0.2$ for the first case and for the second superimposed by external noise with $\sigma = 0.5$. Especially striking is the quadratic behaviour of the estimates for the diffusion coefficient. In the case of a finite sampling rate defined by τ_0 this has already been detected in the approximation (2.7) with the correction term $\tau^2(D^{(1)})^2$. An estimate that is adapted to this correction would reduce, though not eliminate the quadratic effect. This is different for the case of a non-zero σ , i.e. in the presence of measurement noise. Hence, it is essential to connect the results for the estimates with the observations for the conditional moments. Whereas a non-zero offset is an indicator for the presence of external noise, finite- τ effects are directly connected to a nonlinear curvature of $M^{(n)}(x, \tau)$ as function of τ . Checking these two indicators gives essential information about which kind of disturbance is present and which definition of the estimate has to be applied. Note that this is a central result for the application of the presented method of reconstruction.

To take a further step, we next consider a process that is affected by a low sampling rate and external noise likewise. The considerations above suggest a certain hierarchy of both effects. The presence of external noise effectively changes the observable process whereas the value of τ_0 just determines when and at which rate the process is observed. Consequently, we start with the considerations of section 2.3 and reformulate (2.20) and (2.21) as

$$\begin{aligned} M_\sigma^{(1)}(y, \tau) &= \int dx M^{(1)}(x, \tau) f(x|y) + \int dx (x - y) f(x|y) \\ &= m^{(1)}(y, \tau) + \gamma_1(y) \end{aligned} \quad (2.31)$$

and

$$\begin{aligned} M_\sigma^{(2)}(y, \tau) &= \int dx [2(x - y)M^{(1)}(x, \tau) + M^{(2)}(x, \tau)] f(x|y) \\ &\quad + \sigma^2 + \int (x - y)^2 f(x|y) \\ &= m^{(2)}(y, \tau) + \gamma_2(y), \end{aligned} \quad (2.32)$$

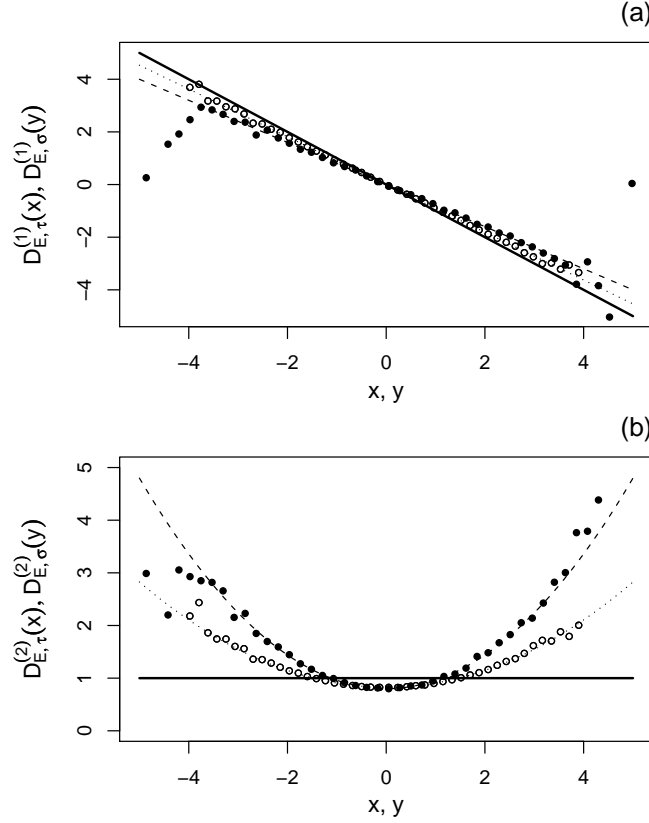


Figure 2.5: Empirical and theoretical estimates $D_{E,\tau}^{(n)}$ and $D_{E,\sigma}^{(n)}$ ($n = 1, 2$) for an Ornstein-Uhlenbeck process ($\alpha = 1, \beta = 1$). The symbols represent the empirical results, the lines the theoretical functions – open symbols and dotted line for $\tau_0 = 0.2$ and $\sigma = 0$ (third-order corrections), full symbols and dashed line for $\tau_0 = 10^{-3}$ and $\sigma = 0.5$. The solid lines indicate the original functions $D^{(1)}(x)$ and $D^{(2)}(x)$.

expanding the moments $M^{(n)}(x, \tau)$ ($n = 1, 2$) according to (2.9) and (2.13). Considering once again an Ornstein-Uhlenbeck process, the conditional moments for the undisturbed process with a finite τ are given by

$$M^{(1)}(x, \tau) = -\tau a(\alpha, \tau)x, \quad (2.33)$$

$$M^{(2)}(x, \tau) = 2\tau[b(\alpha, \beta, \tau) + c(\alpha, \tau)x^2] \quad (2.34)$$

with appropriate definitions of the functions $a = a(\alpha, \tau)$, $b = b(\alpha, \beta, \tau)$ and $c = c(\alpha, \tau)$, dependent on how many terms we consider in (2.13). (The explicit functions for third-order approximations are given in the appendix to this chapter.) Inserting (2.33) and (2.34) into (2.31–2.32), we obtain

$$m^{(1)}(y, \tau) = \tau[-a(\alpha, \tau) (1 - \frac{\sigma^2}{\lambda^2}) y] \quad (2.35)$$

$$m^{(2)}(y, \tau) = 2\tau\{b(\alpha, \beta, \tau) - a(\alpha, \tau) [\frac{\sigma^2 s^2}{\lambda^2} + (\frac{\sigma^4}{\lambda^4} - \frac{\sigma^2}{\lambda^2}) y^2] + c(\alpha, \tau) [\frac{s^2 \sigma^2}{\lambda^2} + \frac{s^4}{\lambda^4} y^2]\}, \quad (2.36)$$

the terms for the offsets $\gamma_1(y)$ and $\gamma_2(y)$ remain unchanged (see (2.27)).

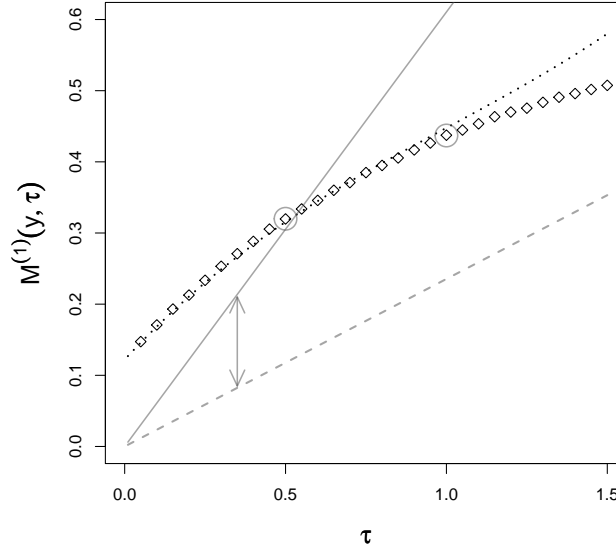


Figure 2.6: Empirical values and theoretical functions for the first conditional moment of an Ornstein-Uhlenbeck process ($\alpha = 1$, $\beta = 1$) superimposed by external measurement noise ($\sigma = 0.5$), and the deviation between the estimate $D_{E,\tau\sigma}^{(1)}(y, \tau_0)$ and the intrinsic function $D^{(1)}(x)$. The empirical values for $M_\sigma^{(1)}(y = -0.6, \tau)$ are given by the symbols, the theoretical function is denoted by the dotted line (third-order approximation). The solid line represents the linear relation $\tau D^{(1)}(y)$, the dashed line the relation $\tau D_{E,\tau\sigma}^{(1)}(y, \tau_0 = 0.5)$ with the estimate according to (2.37). The grey circles indicate the values $M_\sigma^{(1)}(y, \tau_0 = 0.5)$ and $M_\sigma^{(1)}(y, \tau_0 = 1.0)$ used for the derivation of the estimate.

Figure 2.6 shows the empirical values $M_\sigma^{(1)}(y, \tau)$ for an Ornstein-Uhlenbeck process, similarly as in figure 2.4 but over a broader range of τ , together with the theoretical approximation according to (2.35) (third-order approximation for $M^{(1)}(x, \tau)$). Additionally, the linear relation $\tau D^{(1)}(y)$ including the intrinsic drift function and the relation $\tau D_{E,\tau\sigma}^{(1)}(y, \tau_0)$ are depicted where the estimate $D_{E,\tau\sigma}^{(n)}(y, \tau_0)$ ($n = 1, 2$) is defined as follows.

Since we have to deal with a nonlinear behaviour due to a comparatively large τ_0 and an (initially unknown) offset due to the measurement noise, simultaneously, neither the estimate defined by (2.16) nor that of (2.26) are practicable. Instead we suggest to define as a third estimate

$$D_{E,\tau\sigma}^{(n)}(y, \tau_0) := \frac{M_\sigma^{(n)}(y, 2\tau_0) - M_\sigma^{(n)}(y, \tau_0)}{n!\tau_0}, \quad (2.37)$$

where τ_0 is the smallest available time increment referring to the sampling rate of the process. That is, we simulate a linear fit to overcome the non-zero offset but minimize the number of fitted points to reduce the impact of the nonlinear curvature.

We notice that $D_{E,\tau\sigma}^{(n)}(y, \tau_0)$ deviates considerably from the intrinsic function $D^{(n)}(x)$, especially for large values of τ_0 . Without measurement noise it converges to the intrinsic function for $\tau_0 \rightarrow 0$. For $\sigma \neq 0$, however, there is still in this limit an offset that corresponds to the deviation already shown in figure 2.4. The estimate $M_\sigma^{(n)}(y, \tau_0)/\tau_0$, using the definition (2.16), would actually provide results closer to $D^{(n)}(x)$ for certain values of τ_0 . The latter, however, diverges for small τ_0 and is for

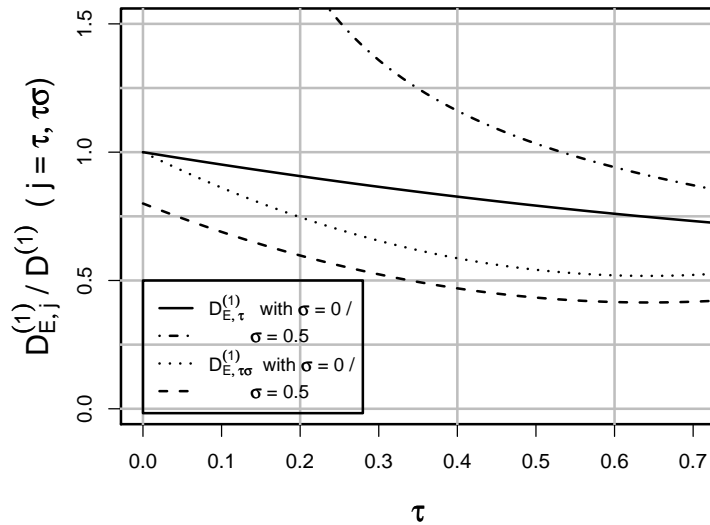


Figure 2.7: Convergence behaviour of the estimates $D_{E,\tau}^{(1)}$ and $D_{E,\tau\sigma}^{(1)}$ for an Ornstein-Uhlenbeck process ($\alpha = 1$, $\beta = 1$) with and without measurement noise ($\sigma = 0.5$). The lines correspond to the theoretical estimates derived on the basis of the third-order approximations for the conditional moments.

this reason not suitable. Figure 2.7 gives an illustration of the discussed behaviour. Note that although the deviations between the estimates and the corresponding intrinsic function are quite large for the most cases, the former can be utilized to extract the actual $D^{(n)}(x)$ with a good precision, as we show in the next section.

2.5 Reconstruction of the intrinsic drift and diffusion functions from the estimates through optimization

In this section, we present how the intrinsic functions $D^{(n)}(x)$ can be reconstructed from the estimates $D_{E,j}^{(n)}$ ($j = \tau, \sigma, \tau\sigma$) applying an optimization scheme. Following [Böttcher et al 2006], where the disturbance by external noise has been considered (but not yet the case of low sampling rates), the basic idea is to minimize the sum of the squared deviations between the empirical estimates and the theoretical correspondents. The general procedure is as follows: For a given data set, the conditional moments are calculated for different values of x or y , respectively, and plotted as function of the respective time increment τ . An offset and/or a nonlinear curvature in these plots are indicators for certain disturbances. The according type of estimate adapted to the respective disturbance is selected, and the empirical values of the estimate are calculated directly from the data. Considering an appropriate parametrization $Q = \{q_i\}$ for the corresponding intrinsic functions $D^{(n)}(x)$, the theoretical estimates are calculated either analytically or numerically. Finally, the sum of the squared deviations between empirical and theoretical estimates is minimized by varying Q and where necessary also the amplitude of the external noise assumed for the respective theoretical estimates. In contrast to [Böttcher et al 2006], the single contributions to the total sum of deviations are additionally weighted by the

variances $\sigma_{D_{E,j}^{(n)}}^2$ of the empirical estimates.

To demonstrate this procedure, we consider the example of a process defined by a linear drift ($D^{(1)}(x) = -ax$) and a quadratic diffusion coefficient ($D^{(2)}(x) = b+cx^2$). A process of this type is observed in various systems ranging from finance to turbulence or physiological time series (cf. [Renner et al 2001a, Wächter et al 2004, Tabar et al 2006]). We have taken the values $a = 1$, $b = 0.1$ and $c = 0.5$ for the parameters and simulated time series consisting of $N = 5 \cdot 10^6$ data points in each case. For the first case, we took a sampling rate of $\tau_0 = 0.2$. A nonlinear curvature but zero offset in the plot of the conditional moments vs. τ suggests to select the estimate $D_{E,\tau}^{(n)}(x)$ as defined in (2.16) and (2.17). We calculated the empirical values for $D_{E,\tau}^{(n)}(x)$ and chose the parametrization $D^{(1)}(x) = -q_1x$ and $D^{(2)}(x) = q_2 + q_3x^2$, $Q = \{q_1, q_2, q_3\}$, to derive the theoretical estimates that will be denoted by $\hat{D}_{E,\tau}^{(n)}(x)$ for the following considerations. This notation indicates that the theoretical estimates are adapted to the empirical ones by variation of the set of parameters Q . Minimizing the sum of weighted quadratic deviations, i.e.

$$\min_Q \left\{ \sum_i \left\{ \frac{1}{\sigma_{D_{E,\tau}^{(1)}}^2} [\hat{D}_{E,\tau}^{(1)}(x_i, \tau_0) - D_{E,\tau}^{(1)}(x_i, \tau_0)]^2 \right. \right. \quad (2.38)$$

$$\left. \left. + \frac{1}{\sigma_{D_{E,\tau}^{(2)}}^2} [\hat{D}_{E,\tau}^{(2)}(x_i, \tau_0) - D_{E,\tau}^{(2)}(x_i, \tau_0)]^2 \right\} \right\},$$

we obtained the optimized reconstructed values $q_1 = 1.0 (\pm 0.01)$, $q_2 = 0.10 (\pm 0.01)$ and $q_3 = 0.51 (\pm 0.01)$ that are in good agreement with the values of the original parameters a , b and c .

For the second case, we took a sampling rate of $\tau_0 = 10^{-3}$ and added measurement noise with an amplitude of $\sigma = 0.25$ to the simulated time series. (This is the example that has already been considered in [Böttcher et al 2006].) We used the parametrization given above and performed the optimization

$$\min_{Q, \tilde{\sigma}} \left\{ \sum_{n=1,2} \sum_i \left\{ \frac{1}{\sigma_{D_{E,\sigma}^{(n)}}^2} [\hat{D}_{E,\sigma}^{(n)}(y_i) - D_{E,\sigma}^{(n)}(y_i)]^2 \right. \right. \quad (2.39)$$

$$\left. \left. + \frac{1}{\sigma_{\gamma_n}^2} [\hat{\gamma}_n(y_i) - \gamma_n(y_i)]^2 \right\} \right\},$$

by varying the parameters Q and $\tilde{\sigma}$, where $\tilde{\sigma}$ denotes the varied value for σ . We obtained the results $q_1 = 0.96 (\pm 0.01)$, $q_2 = 0.10 (\pm 0.01)$, $q_3 = 0.51 (\pm 0.01)$ and $\tilde{\sigma} = 0.25 (\pm 0.01)$.

For the third and last case, we took a sampling rate of $\tau_0 = 0.2$ as in the first case but added measurement noise with an amplitude of $\sigma = 0.25$ to the simulated time series. We calculated the empirical estimates as defined in (2.37) and compared them with the theoretical values derived according to

$$\hat{D}_{E,\tau\sigma}^{(n)}(y, \tau_0) := \frac{m^{(n)}(y, 2\tau_0) - m^{(n)}(y, \tau_0)}{n!\tau_0}. \quad (2.40)$$

The (extrapolated) offsets of the corresponding conditional moments are given by

$$\gamma_{n,E}(y, \tau_0) := 2M_\sigma^{(n)}(y, \tau_0) - M_\sigma^{(n)}(y, 2\tau_0) \quad (2.41)$$

and

$$\hat{\gamma}_{n,E}(y, \tau_0) := 2m^{(n)}(y, \tau_0) - m^{(n)}(y, 2\tau_0) + \hat{\gamma}_n \quad (2.42)$$

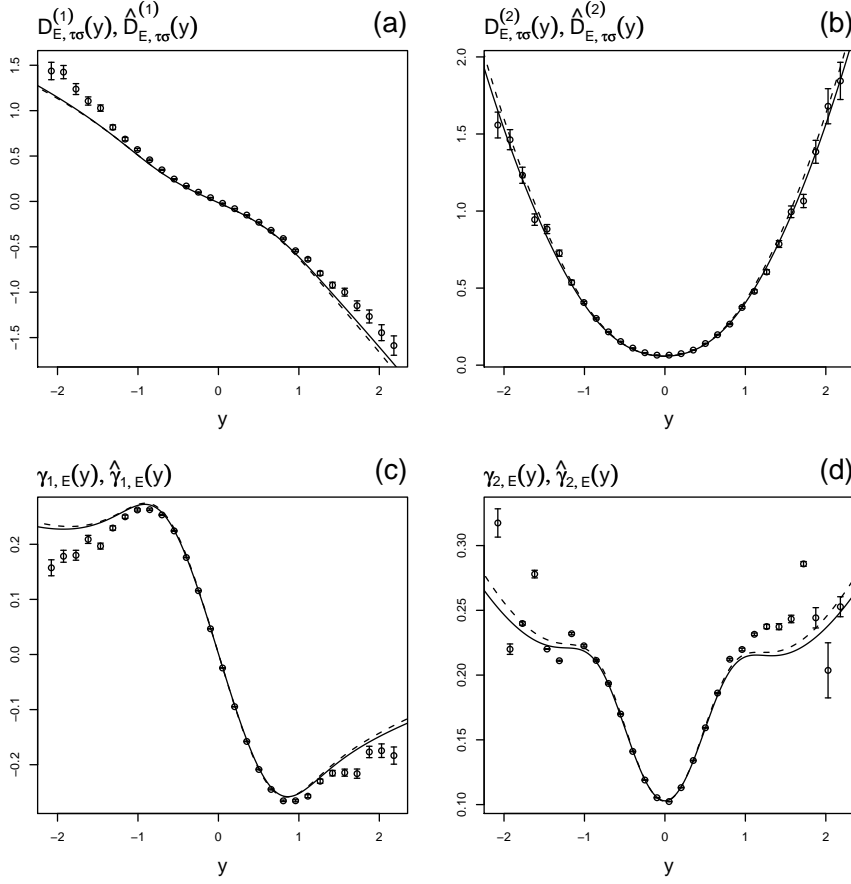


Figure 2.8: Empirical (symbols) and theoretical (solid and dashed lines) estimates $D_{E, \tau \sigma}^{(n)}$ and extrapolated offsets $\gamma_{n, E}$ ($n = 1, 2$) for a process defined by the intrinsic functions $D^{(1)}(x) = -ax$ ($a = 1$) and $D^{(2)}(x) = b + cx^2$ ($b = 0.1$, $c = 0.5$) sampled with $\tau_0 = 0.2$ and spoiled by external measurement noise ($\sigma = 0.25$). Solid lines correspond to reconstructed and optimized set of parameters (Q , $\tilde{\sigma}$) and dashed lines to original parameters (a , b , c , σ).

for the empirical and the theoretical case, respectively. We performed the optimization task

$$\min_{Q, \tilde{\sigma}} \left\{ \sum_{n=1,2} \sum_i \left\{ \frac{1}{\sigma_{D_{E, \tau \sigma}^{(n)}}^2} [\hat{D}_{E, \tau \sigma}^{(n)}(y_i, \tau_0) - D_{E, \tau \sigma}^{(n)}(y_i, \tau_0)]^2 + \frac{1}{\sigma_{\gamma_{n, E}}^2} [\hat{\gamma}_{n, E}(y_i, \tau_0) - \gamma_{n, E}(y_i, \tau_0)]^2 \right\} \right\} \quad (2.43)$$

by varying Q and $\tilde{\sigma}$, and obtained as results $q_1 = 0.98 (\pm 0.01)$, $q_2 = 0.10 (\pm 0.01)$, $q_3 = 0.47 (\pm 0.01)$ and $\tilde{\sigma} = 0.25 (\pm 0.01)$. Figures 2.8(a–d) shows, exemplarily for this last case, the empirical values for the drift and diffusion estimates and the extrapolated offsets in comparison to the theoretical ones derived for the parameter sets $(Q, \tilde{\sigma})$ and (a, b, c, σ) .

2.6 Handling the estimates and a note on the robustness of fixed points

For real data sets, an optimization procedure to reconstruct the intrinsic drift and diffusion functions from the disturbed estimates, as presented in the preceding section, is not always realizable. Selecting a parametrization with a too high number of single parameters or considering too many orders of τ for the conditional moments, the procedure rapidly becomes very complex and expensive.

Furthermore, the procedure is strictly speaking not applicable to processes that are not Langevin or Markov processes, respectively. As a rule, the data set at hand should be tested with regard to the Markov property prior to the analysis. Often one can find a typical length scale above that the process is Markovian even when this property is violated below (see e.g. [Renner et al 2001b]). Promising is, in this connection, also the fact that the presence of measurement noise itself spoils the Markov properties of an underlying Markov process, as shown in [Kleinhans et al 2007]. Hence, limitations at small scales must not stem from the intrinsic properties of the underlying process dynamics but might be the result of artificial noise sources that can be controlled much better than the actual intrinsic dynamics.

As stated in the introduction, it might moreover even for non-Markovian processes make sense to estimate at least the drift coefficient of such a process. The reconstruction of the drift coefficient is not necessarily affected by a non-Markovianity of the process. Against this, we have seen in sections 2.2–2.4 that it is significantly disturbed by finite- τ and noise effects. It arises the question if the process dynamics can be reduced to a minimal characteristic behaviour of the process that is not or only weakly influenced by these effects. Reducing the drift dynamics to some characteristic behaviour leads to the definition of its fixed points. Hence, the question is if and how the fixed points of a certain process are affected by the discussed resolution and noise effects. To give an answer, we assume for the following considerations again that we deal with a Langevin process.

From the analytical results obtained above we find that for an Ornstein-Uhlenbeck process the position of the system's fixed point is not at all influenced by the discussed effects. The Ornstein-Uhlenbeck process is an absolutely symmetric process and the robustness of its fixed point can just as well be deduced from simple symmetry arguments. For non-symmetric processes, the case is more complicated. In figures 2.9(a–d), we show the theoretical estimates of the drift coefficients for four exemplary processes in comparison to the respective intrinsic functions $D^{(1)}(x)$ (each time as solid black line). The estimates for the finitely resolved process and that superimposed by external noise – given by the dotted and dashed lines, resp. – have been derived according to (2.16) and (2.26). For figure 2.9(a), we have considered a process with linear drift and quadratic diffusion coefficient, both symmetric with respect to their fixed point. The estimates seem to reproduce the fixed point robustly. (A very small deviation stemming from the third-order finite- τ correction is not resolved by the figure.) Figure 2.9(b) also shows a symmetric process but with a multiple stability. The central fixed point is robust, whereas the two outer move with the estimates. It is clear that the central fixed point is determined by another symmetry than the two outer ones. While the shape of the drift coefficient is compressed or stretched due to the finite- τ as well as measurement-noise effects which strongly affects the outer fixed points, the central fixed point remains unchanged. The latter is seen here as the centre of symmetry. Figure 2.9(c) shows a process with asymmetric drift and constant diffusion, and figure 2.9(d) one that is characterized by a shifted diffusion coefficient, otherwise it is the same process as in figure 2.9(a). In both cases, the estimates differ from the intrinsic function for

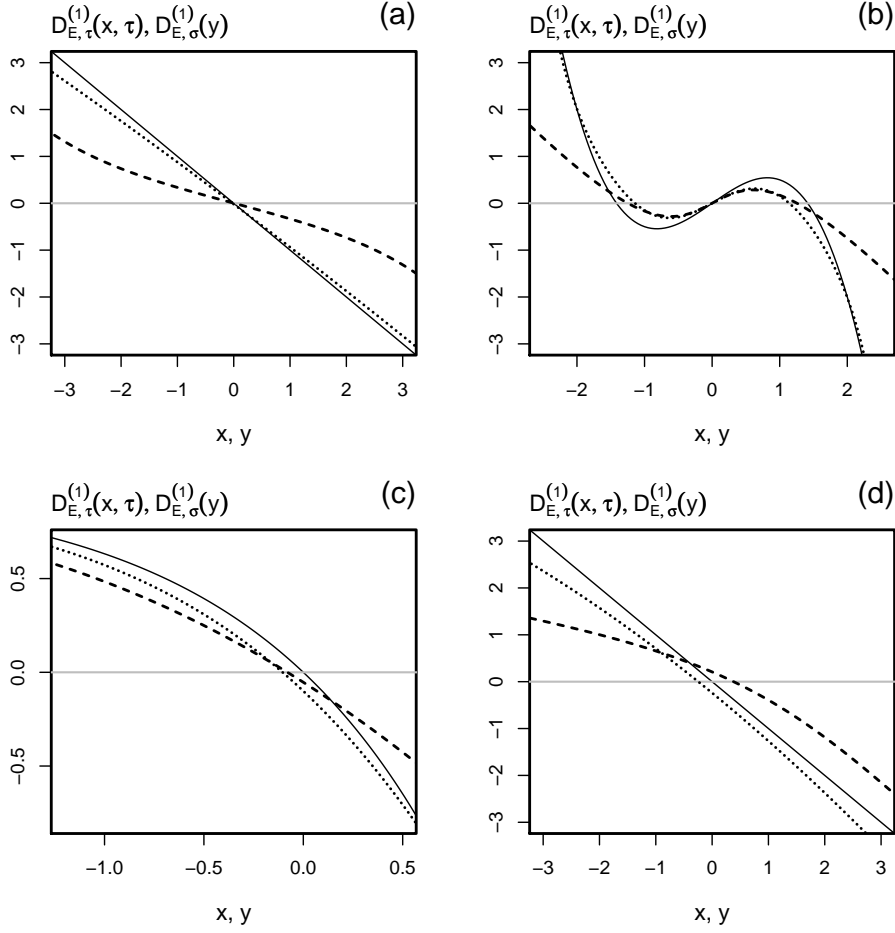


Figure 2.9: Theoretical drift estimates for four exemplary processes. (a) $D^{(1)}(x) = -ax$ and $D^{(2)}(x) = b + cx^2$ ($a = 1, b = 1, c = 0.5$), (b) $D^{(1)}(x) = ax - bx^3$ and $D^{(2)}(x) = c$ ($a = 1, b = 0.5, c = 1$), (c) $D^{(1)}(x) = -\exp(x) + 1$ and $D^{(2)}(x) = b + cx^2$ ($a = 1, b = 1, c = 0.5$), (d) $D^{(1)}(x) = -ax$ and $D^{(2)}(x) = b + c(x + d)^2$ ($a = 1, b = 1, c = 0.5, d = 2$) – each time with $\tau_0 = 0.2, \sigma = 0$ (dotted) and $\tau_0 \ll 1, \sigma = 1$ (dashed).

$D^{(1)}(x)$ to a certain extent, and the fixed point is shifted due to the asymmetries.

In summary, we can conclude that symmetries in $D^{(1)}(x)$ and $D^{(2)}(x)$ are of extreme importance for the robustness of fixed points in the presence of measurement noise and for finitely resolved processes. Asymmetries in the estimates for drift and/or diffusion coefficients are therefore to be utilized as indicators and warnings for potential shiftings of the fixed points. Finite- τ and measurement noise effects may lead to similar, i.e. equally directed, but also to opposite deviations of the estimated fixed points compared to the intrinsic ones. Note that the chosen values for σ correspond to very strong measurement noise. Similarly, the chosen value for τ_0 is quite large comparing it with the intrinsic time of the process that is given by a^{-1} for a process with linear drift. The deviations between estimates and intrinsic functions are accordingly large and should be regarded as upper limits rather than as typical results. Nevertheless, the presented results show that even for these cases the structure of the considered processes, i.e. their approximate functional behaviour, could be unveiled by the estimates. A refined approach is to relate the

amplitude of the external noise as well as the resolution of the sampled time series with the nonlinearities of the intrinsic drift and diffusion functions. Since the fixed-point analysis is a local linear method, it is affected if and only if the disturbance brings the dynamics out of the linear vicinity of the fixed point.

2.7 Conclusions

We have investigated how two different aspects influence the reconstruction of Langevin processes – a finite sampling rate and the presence of external measurement noise. For both disturbances we identified indicators and introduced corresponding estimates for the drift and diffusion functions defining the process. The different estimates essentially differ in how the limit $\tau \rightarrow 0$ for the reconstruction is realized (cf. the functional relation in (2.3)). Each kind of disturbance complicates this limit in a specific way and the proposed estimates are adapted to the particular difficulty. The specific relevance of the presented achievements is that the two disturbances are not only discussed separately, as in [Friedrich et al 2002], [Ragwitz & Kantz 2001] and [Böttcher et al 2006], but also in a combined way. This allows for a more advanced application of the method of reconstruction to real data. Note that our method also covers the case of strong noise.

We have presented two possibilities of technical application – one is to reconstruct the intrinsic drift and diffusion functions from the estimates utilizing an optimization scheme, and the other is to reduce the reconstruction of the process dynamics to the estimation of its fixed points. The proposed optimization scheme for the reconstruction can be seen as an alternative to the iterative procedure introduced in [Kleinhans et al 2005] with the advantage that it can be applied to processes suffering from a finite sampling rate and the spoiling by external noise likewise. For the reconstruction of the fixed points we have worked out the relevance of the symmetries of the underlying dynamical equations. For asymmetric processes we showed that the derived estimates indicate fixed points that are shifted in comparison to the intrinsic functions due to a finite sampling rate and/or the presence of external measurement noise. Thus, we obtained an indicator for the robustness and the quality of estimated fixed points and at the same time defined a frame for the application of a reliable fixed-point method. Likewise, the disturbances are to be related to the nonlinearities of the intrinsic functions. With this information, that is just obtained from the proposed method of data analysis, a model for the considered system can be put up and fine-tuned optimizing the estimates by iterative steps.

2.8 Appendix: Derivation of third-order terms

As third-order approximations for the conditional moments we find

$$\begin{aligned}
M^{(1)}(x, \tau) &= \tau D^{(1)} + \frac{\tau^2}{2} \left[D^{(1)}(D^{(1)})' + D^{(2)}(D^{(1)})'' \right] \\
&+ \frac{\tau^3}{6} \left[D^{(1)}(D^{(1)})'(D^{(1)})' + D^{(1)}D^{(1)}(D^{(1)})'' \right. \\
&\quad + 3D^{(2)}(D^{(1)})'(D^{(2)})'' + 2D^{(1)}D^{(2)}(D^{(1)})''' \\
&\quad + D^{(1)}(D^{(2)})'(D^{(1)})'' + D^{(2)}(D^{(1)})''(D^{(2)})'' \\
&\quad \left. + 2D^{(2)}(D^{(2)})'(D^{(1)})''' + D^{(2)}D^{(2)}(D^{(1)})'''' \right] \\
&+ \mathcal{O}(\tau^4), \tag{2.44}
\end{aligned}$$

and similarly for the second conditional moment

$$\begin{aligned}
M^{(2)}(x, \tau) &= 2\tau D^{(2)} + \tau^2 \left[(D^{(1)})^2 + 2D^{(2)}(D^{(1)})' + D^{(1)}(D^{(2)})' + D^{(2)}(D^{(2)})'' \right] \\
&+ \frac{\tau^3}{3} \left[3D^{(1)}D^{(1)}(D^{(1)})' + 4D^{(2)}(D^{(1)})'(D^{(1)})' \right. \\
&\quad + 7D^{(1)}D^{(2)}(D^{(1)})'' + 3D^{(1)}(D^{(1)})'(D^{(2)})' \\
&\quad + 4D^{(2)}D^{(2)}(D^{(1)})''' + 7D^{(2)}(D^{(2)})'(D^{(1)})'' \\
&\quad + 4D^{(2)}(D^{(1)})'(D^{(2)})'' + D^{(1)}D^{(1)}(D^{(2)})'' \\
&\quad + 2D^{(1)}D^{(2)}(D^{(2)})''' + D^{(1)}(D^{(2)})'(D^{(2)})'' \\
&\quad + D^{(2)}(D^{(2)})''(D^{(2)})'' + 2D^{(2)}(D^{(2)})'(D^{(2)})''' \\
&\quad \left. + D^{(2)}D^{(2)}(D^{(2)})'''' \right] + \mathcal{O}(\tau^4). \tag{2.45}
\end{aligned}$$

For an Ornstein-Uhlenbeck process ($D^{(1)}(x) = -\alpha x$ and $D^{(2)}(x) = \beta$) this gives

$$M^{(1)}(x, \tau) = -\tau\alpha x + \frac{\tau^2\alpha^2 x}{2} - \frac{\tau^3\alpha^3 x}{6} + \mathcal{O}(\tau^4), \tag{2.46}$$

$$M^{(2)}(x, \tau) = 2\beta\tau + \tau^2(\alpha^2 x^2 - 2\alpha\beta) + \frac{\tau^3}{3}(-3\alpha^3 x^2 + 4\alpha^2\beta) + \mathcal{O}(\tau^4). \tag{2.47}$$

Hence, the functions $a = a(\alpha, \tau)$, $b = b(\alpha, \beta, \tau)$ and $c = c(\alpha, \tau)$, introduced in section 2.4, are given by

$$a(\alpha, \tau) = \alpha - \frac{\tau\alpha^2}{2} + \frac{\tau^2\alpha^3}{6}, \tag{2.48}$$

$$b(\alpha, \beta, \tau) = \beta - \tau\alpha\beta + \frac{2\tau^2}{3}\alpha^2\beta \tag{2.49}$$

and

$$c(\alpha, \tau) = \frac{\tau}{2}\alpha^2 - \frac{\tau^2}{2}\alpha^3. \tag{2.50}$$

Bibliography

- [Anahua et al 2008] E. Anahua, St. Barth, and J. Peinke: Markovian Power Curves for Wind Turbines. In: *Wind Energy* 11(3), 219–232 (2008).
- [Böttcher et al 2006] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase: Reconstruction of Complex Dynamical Systems Affected by Strong Measurement Noise. In: *Phys. Rev. Lett.* 97, 090603 (2006).
- [Friedrich & Peinke 1997] R. Friedrich and J. Peinke: Statistical properties of a turbulent cascade. In: *Phys. Rev. Lett.* 78, 863–866 (1997).
- [Friedrich et al 2000] R. Friedrich, S. Siegert, J. Peinke, St. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, G. and Pfister: Extracting model equations from experimental data. In: *Phys. Lett. A* 271, 217–222 (2000).
- [Friedrich et al 2002] R. Friedrich, Ch. Renner, M. Siefert, and J. Peinke: Comment on “Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data”. In: *Phys. Rev. Lett.* 89, 149401 (2002).
- [Gottschall & Peinke 2008] J. Gottschall and J. Peinke: How to improve the estimation of power curves for wind turbines. In: *Environ. Res. Lett.* 3, 015005 (2008).
- [Kern et al 2005] M. Kern, O. Buser, J. Peinke, M. Siefert and L. Vulliet: Stochastic analysis of single particle segregational dynamics. In: *Phys. Lett. A* 336, 428–433 (2005).
- [Kleinhans et al 2005] D. Kleinhans, R. Friedrich, A. Nawroth, and J. Peinke: An iterative procedure for the estimation of drift and diffusion coefficients of Langevin processes. In: *Phys. Lett. A* 346, 42–46 (2005).
- [Kleinhans et al 2007] D. Kleinhans, R. Friedrich, M. Wächter, and J. Peinke: Markov properties in presence of measurement noise. In: *Phys. Rev. E* 76, 041109 (2007).
- [Kleinhans & Friedrich 2007] D. Kleinhans and R. Friedrich: Quantitative Estimation of Drift and Diffusion Functions from Time Series Data. In: *Wind Energy – Proc. of the Euromech Colloquium* ed J. Peinke, P. Schaumann, St. and Barth; Springer, Berlin, 129–134 (2007).
- [Kriso et al 2002] S. Kriso, J. Peinke, R. Friedrich and P. Wagner: Reconstruction of dynamical equations for traffic flow. In: *Phys. Lett. A* 299, 287–291 (2002).
- [Ragwitz & Kantz 2001] M. Ragwitz and H. Kantz: Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data. In: *Phys. Rev. Lett.* 87, 254501 (2001).
- [Renner et al 2001a] Ch. Renner, J. Peinke, and R. Friedrich: Experimental indications for Markov properties of small-scale turbulence. In: *J. Fluid. Mech.* 433, 383–409 (2001).
- [Renner et al 2001b] Ch. Renner, J. Peinke, and R. Friedrich: Evidence of Markov properties of high frequency exchange rate data. In: *Physica A* 298, 499–520 (2001).
- [Risken 1989] H. Risken: *The Fokker-Planck Equation*. Springer, New York (1989).
- [Ryskin 1997] G. Ryskin: Simple procedure for correcting equations of evolution – Application to Markov processes. In: *Phys. Rev. E* 56, 5123–5127 (1997).

- [Siefert et al 2003] M. Siefert, A. Kittel, R. Friedrich, and J. Peinke: On a quantitative method to analyze dynamical and measurement noise. In: *Europhys. Lett.* 61, 466–472 (2003).
- [Siefert & Peinke 2004] M. Siefert and J. Peinke: Reconstruction of the deterministic dynamics of stochastic systems. In: *Int. J. Bifurcation Chaos Appl. Sci. Eng.* 14, 2005–2009 (2004).
- [Siegert et al 1998] S. Siegert, R. Friedrich, and J. Peinke: Analysis of data sets of stochastic systems. In: *Phys. Lett. A* 271, 275–280 (1998).
- [Sura & Barsugli 2002] P. Sura and J. Barsugli: A note on estimating drift and diffusion parameters from timeseries. In: *Phys. Lett. A* 305, 304–311 (2002).
- [Tabar et al 2006] M. R. R. Tabar, F. Ghasemi, J. Peinke, R. Friedrich, K. Kaviani, F. Taghavi, S. Sadeghi, G. Bizhani, and M. Sahimi: New computational approaches to analysis of interbeat intervals in human subjects. In: *Computing in Science and Engineering* 8(2), 54–65 (2006).
- [Wächter et al 2004] M. Wächter, F. Riess, T. Schimmel, U. Wendt, and J. Peinke: Stochastic analysis of different rough surfaces. In: *Eur. Phys. J. B* 41, 259–277 (2004).

Chapter 3

A phenomenological model for the power performance of wind turbines

The work that is presented in this chapter has arisen for the most part as subproject of the cooperative project “Wind turbulences and their impact on the utilization of wind energy” that was supported and financed by the BMBF (Federal Ministry of Education and Research) during 2005 and 2008. This part of the project deals with the analysis and modelling of the power performance of a wind turbine and especially the influence of turbulent wind structures on it.

After a more general introduction to power performance testing, the characterization of small-scale structures in wind speed time series is discussed in some detail. In an analogous analysis of respective power output data, it is shown how the effects of turbulence appear in the power performance of a wind turbine. In particular, a dynamical modelling of the power performance based on high-frequency data is introduced, describing the response dynamics of the turbine to the wind fluctuations as a set of Langevin equations. This stochastic modelling enables a quite flexible treatment of the turbulent small-scale structures. The respective methods for reconstructing the underlying dynamical equations in terms of drift and diffusion, introduced in Chapter 1 and 2, are applied to define a dynamical power characteristic that is discussed in detail with respect to its applicability of reproducing the actual power performance dynamics of an investigated wind turbine. Two appendices include a previously published paper as well as a detailed description of the data that were measured within the scope of the above mentioned project and primarily used for the present analyses.

3.1 Basic considerations on power performance testing

3.1.1 Power performance and effective control

Wind turbines are mechanical devices that are specifically designed to convert a part of the kinetic energy of a wind field into useful mechanical energy. Since nowadays almost all wind turbines are built to convert this mechanical energy in a second step into electricity, the term wind turbine often also includes the power generator unit. This is the definition I refer to. Furthermore, I restrict my considerations to large-scale horizontal-axis wind turbines and assume that the produced electrical energy is directly fed into the grid. This is not an essential restriction but facilitates putting the presented work in a certain context. For a detailed overview of different types of wind turbines and corresponding modes of operation see e.g. [Burton et al 2001].

A common synonym for the term wind turbine is the more technical denotation wind energy converter system (WECS), summarizing similarly the single components of the rotor, the transmission system and the power generator unit. Interpreting a wind turbine as a complex system, it is dealt with a complex structure of interacting subsystems that transfers the fluctuating wind field as input variable into corresponding fluctuations of electrical power. The general relation between a representative wind velocity and the output of electrical power is covered by the term power performance. The central aim of power performance testing is to quantify this relation for a specific wind turbine in order to characterize, certify or monitor its behaviour. It is obvious that a key issue of power performance testing is the measurement of the wind velocity. In principle, one is interested in the free-field velocity, i.e. the velocity in the rotor plane if the wind turbine was removed. Since the wind turbine however distorts the wind field, a corresponding measurement in the rotor plane or nearby is not of use – at least not without further corrections. Instead the upstream velocity is usually chosen as representative of the free-field velocity, and measured with a meteorological mast in a certain distance in front of the turbine. A second approximation consists in considering not the entire field acting on the rotor but reducing it to the horizontal component of the three-dimensional field on the height of the turbine's hub. The consideration on how representative this wind speed actually is for the wind field that drives the rotor is taken up in the discussion in section 3.4.

Since the early 1980s, several groups of experts developed recommendations for defining and determining the power performance of a wind turbine. Continuously developed, they were adopted in a guideline by the International Electrotechnical Commission (IEC) as the international standard IEC 61400-12 and the revised version IEC 61400-12-1 [IEC 2005a]. Following these common guidelines for power performance testing, the power performance characteristic of a wind turbine is defined by its measured power curve and the corresponding estimated annual energy production (AEP). The measured power curve is determined by collecting simultaneous measurements of wind speed, as defined above, and power output at the test site for a period that is long enough to establish a statistically significant database over a range of wind speeds and under varying wind and atmospheric conditions. The power is, at this, defined as the net power, i.e. the produced electrical power less all power losses caused by the turbine's internal consumption. Subsequently to the data collection, a functional relation, i.e. a two-dimensional curve of power output vs. wind speed, is extracted out of this data. The AEP is derived by applying the measured power curve to a reference distribution of wind speed for the test site and assuming a specified availability of the wind turbine.

The general shape of a power curve is deduced from the relation

$$P(u) = C_P(u)P_u(u) \quad \text{with} \quad P_u(u) = \frac{1}{2}\rho_0 Au^3, \quad (3.1)$$

where P_u is the power in the wind passing with a speed u through the rotor area A at a density ρ_0 , and P is the extracted power – i.e. a power curve is roughly characterized by a cubic increase of the power output with the wind speed. The functional behaviour of the power coefficient $C_P(u)$ is the result of certain control strategies as well as of physical limits. A theoretical limit for the maximum extractable power is given by the Betz limit $C_{P,\max} = 16/27 \approx 0.59$ that is derived on the basis of momentum theory, modelling the wind turbine's rotor as an actuator disc. Although it is based on a strongly simplified approach, the Betz limit is a widely used and accepted value. The power coefficients of modern commercial wind turbines reach values of about 0.45 and are thus considerably below this theoretical limit and especially not constant for the whole range of wind speeds. Physical aspects that limit the value of the power coefficient are e.g. the finite number of blades and losses due to the drag and stall effects of the blades [Burton et al 2001, Bianchi et al 2006].

Beside extracting as much energy as possible from the wind, the most essential control objective for a WECS is to limit the generated power to a specified value in order to avoid overloading under high wind conditions. This value is called rated power and is a basic design parameter of a wind turbine. The wind speed at that the turbine reaches this value is referred to as rated wind speed. Furthermore, a power curve is defined by the values of cut-in and cut-out wind speed that specify the range of operation.

An ideal power curve, that is characterized by a cubic dependency of the power on the wind speed according to (3.1) until the value of rated wind speed and then shows a constant line at rated power, corresponds to a $C_P(u)$ -curve with a constant line at $C_{P,\max}$ for the range of wind speeds between cut-in and rated value that is followed by a decrease of $C_P(u)$ proportional to u^{-3} until the wind speed reaches the cut-out value. The shape of real power curves typically deviates from this ideal behaviour which is at least to some extent the result of more refined control strategies. In addition to maximizing the energy capture of a turbine and assuring its safe operation, important control objectives are to prevent the system from excessive mechanical loads as well as to maintain a certain power quality of the generated power that is fed into the grid. The power curve that is determined as the result of a power performance measurement eventually provides an insight into the effective control behaviour of the wind turbine, disregarding everything that happens between the recording of the wind speed and the output of power by the wind turbine. It can be understood as a summarizing characteristic for the entire WECS.

3.1.2 Standard procedure for power performance testing (IEC 61400-12-1)

In this section, I briefly summarize the procedure for measuring the power performance characteristic of a wind turbine as specified by the international standard IEC 61400-12-1 (cf. [IEC 2005a]). The purpose of this standard is to provide a uniform methodology ensuring consistency, accuracy and reproducibility in the measurement and analysis of the power performance of wind turbines. It clearly should be seen as a compromise solution constituting the minimum requirements of a power performance measurement, instead of as the final solution, and a procedure of data analysis that is applicable without specific background knowledge.

The first part of the standard gives a detailed description of the necessary preparations for the performance test. This contains the criteria the test equipment has

to fulfil as well the specifications for the location and setup of the meteorological mast that is used to measure the wind speed and further meteorological parameters as the wind direction, the temperature and air pressure. It also defines how to choose the measurement sector, i.e. the range of wind directions that are considered to be valid for a representative measurement of wind speed. Thereby, those wind directions are excluded for that the mast is in the wake of the turbine. However, it is not ensured that a certain part of a wind reaches the mast before the turbine, assuming something like a frozen structure in its main direction of flow.

A more refined assessment of the terrain at the test site is the aim of the optional site calibration procedure that accounts for further significant obstacles beside the turbine itself. If a first investigation of the terrain yields that an enlarged site calibration is necessary, a table of flow correction factors is determined in an additional measurement and these factors are applied to the measured wind speed to reduce the effect of terrain complexities on the power performance measurement.

The actual measurement procedure consists in collecting the data for the different variables that should meet a set of further clearly defined criteria to ensure that the data are of sufficient quantity and quality to determine the power performance characteristics of the wind turbine accurately. The selected data is averaged over periods of 10 min, and these mean values together with the corresponding standard deviations are used for the analysis. Dependent on the type of turbine, either the wind speed averages (for turbines with active power control) or the mean values of power output (for stall-regulated turbines) are normalized to a reference air density.

The measured power curve is determined on the basis of the normalized values by applying the so-called method of bins. That is, the wind speed values are divided into intervals of a width of 0.5 m/s each, and for each interval bin averages of wind speed and power output are calculated according to

$$V_i = \frac{1}{N_i} \sum_{j=1}^{N_i} V_{n,i,j} \quad \text{and} \quad P_i = \frac{1}{N_i} \sum_{j=1}^{N_i} P_{n,i,j} \quad (3.2)$$

where $V_{n,i,j}$ and $P_{n,i,j}$ are the normalized values of wind speed and power averaged over 10 min, V_i and P_i the corresponding bin averages, and N_i is the number of 10 min data sets in the i th bin. The power curve is assumed to be complete or reliable if each bin includes a minimum of 30 min of sampled data and the entire database covers a minimum period of 180 h of data sampling. The covered wind speed range shall at least extend from 1 m/s below cut-in wind speed to 1.5 times the wind speed at 85 % of the rated power of the wind turbine.

Subsequently, the AEP (annual energy production) is estimated by applying the measured power curve to a reference wind speed distribution for different average values according to

$$\text{AEP} = N_h \sum_{i=1}^N [F(V_i) - F(V_{i-1})] \left(\frac{P_{i+1} - P_i}{2} \right). \quad (3.3)$$

N_h is the number of hours in one year, V_i and P_i are the points of the measured power curve for all N bins and $F(V)$ is the specified theoretical wind speed distribution. It is established to assume for this purpose a Rayleigh distribution that is defined according to

$$F(V) = 1 - \exp \left[-\frac{\pi}{4} \left(\frac{V}{V_{\text{ave}}} \right)^2 \right] \quad (3.4)$$

where V_{ave} is the considered average wind speed at hub height.

The wind speed and power values of the measured power curve define the power coefficient C_P by

$$C_{P,i} = \frac{P_i}{\frac{1}{2} \rho_0 A V_i^3} \quad (3.5)$$

with ρ_0 as the reference air density and A as the area swept by the wind turbine's rotor, that is presented as a function of the averaged wind speed values V_i for each bin.

The standard, furthermore, provides a detailed description of the evaluation of uncertainty in the power performance measurement. Single uncertainty components are divided into category A and category B uncertainties. For the first category, the magnitude of the respective uncertainties is deduced from the measurement as e.g. the standard deviation of the distribution of normalized power data. The second category of uncertainties is related to the instruments, the data acquisition system and the terrain surrounding the power performance test site, and its magnitudes are given as characteristic values.

An extension of the presented procedure is given by the standard IEC 61400-12-2 that is currently under consideration. This complement gives a detailed description of evaluating power performance on the basis of nacelle anemometry. Due to a very complex terrain it is sometimes not possible to find a suitable location for the meteorological mast. Instead the anemometer is mounted directly on or near the test turbine's nacelle. The problem is however in this case that the measured wind speed is strongly affected by the rotor of the test turbine. The revised standard therefore provides methods for determining and applying appropriate corrections in terms of a so-called nacelle wind speed transfer function (NTF).

3.1.3 Averaged power performance versus short-term dynamics

A considerable drawback of the standard procedure for power performance testing, that is presented above, is the quite long measurement period that is necessary to obtain statistically significant results, i.e. essentially to reach statistical convergence, on the basis of the mean values of wind speed and power averaged over periods of 10 min. In addition, this averaging accounts for systematic errors in the final results as shall be shown with a simple calculation (see e.g. [Risø-M-2632]). Expressing the wind speed as $u = \bar{u} + u'$, where \bar{u} is the mean value with respect to a specified time period, i.e. $\bar{u} \equiv \langle u \rangle$ as an average for a set of discrete points, and u' are the high-frequency fluctuations around this mean value, it is straightforward to show that $\langle P(u) \rangle \neq P(\bar{u})$ holds for a nonlinear relation $P(\bar{u})$ and a non-vanishing variance of the fluctuations u' . Therefore, the function $P(u)$ is firstly expanded in the Taylor series

$$P(u) = P(\bar{u}) + \frac{\partial P(\bar{u})}{\partial u} u' + \frac{1}{2} \frac{\partial^2 P(\bar{u})}{\partial u^2} u'^2 + \mathcal{O}(u'^3). \quad (3.6)$$

Assuming that the average of the fluctuations u' vanishes, i.e. $\langle u' \rangle = 0$, and neglecting the terms $\mathcal{O}(u'^3)$, one obtains for the averaged power values

$$\langle P(u) \rangle = P(\bar{u}) + \frac{1}{2} \frac{\partial^2 P(\bar{u})}{\partial u^2} \sigma^2 \quad (3.7)$$

with $\sigma^2 \equiv \langle u'^2 \rangle$. According to this result, it follows that fluctuations in the wind speed, that are broadly referred to as turbulence, together with the nonlinearities of the power curve give rise to systematic errors in the averaged power values. (A more detailed characterization of the term turbulence is given in the next following section.) Introducing the turbulence intensity $I_u \equiv \sigma/\bar{u}$, that is commonly utilized as a site-specific quantity, the corresponding deviations are proportional to I_u^2 . The obtained deviations are often related to so-called turbulence effects, holding the turbulent wind fluctuations responsible for them. In several approaches, it has been attempted to incorporate these deviations as correction terms into an advanced procedure of power performance analysis

[Risø-M-2632, Albers & Hinsch 1996, Kaiser et al 2007, Albers et al 2007]. A basic limitation is however at this that the exact relation $P(\bar{u})$ is in principle unknown.

An alternative approach is to consider for the analysis of power performance not only the averaged values over 10 min but also to utilize the information from smaller time scales – either by introducing shorter averaging periods or by using directly the highly sampled data. At this, however, another difficulty arises that is referred to as lack-of-correlation problem (see e.g. [Risø-M-2632]). On small time scales the fluctuations of wind speed and power output are far from perfectly correlated due to several reasons. Firstly, the dynamics of the wind field is averaged over the area that is swept by the rotor before it is reflected in the power fluctuations. That is, the rotor acts as a low-pass filter on the wind speed fluctuations. The averaging time T applied to the data must be large enough not to feel the respective influence if this effect is not considered in an appropriate dynamical description. The same holds for the control dynamics of the wind turbine. Adjusting the control strategy to the varying wind conditions, the WECS shows a certain response behaviour that possibly delays or integrates over the occurring fluctuations. In [Risø-M-2632] it is argued that the corresponding dynamic time constant of the WECS will start to play a role for the evaluation of power performance if its order of magnitude becomes larger than about $0.1T$, where T is the averaging time defined above. A further relevant issue is the spatial distance between the wind speed measurement and the position of the turbine. This distance, that is usually between two and four times the rotor diameter according to [IEC 2005a], introduces an additional delay in the fluctuations and a corresponding lack of coherence. In this connection, it also becomes important what kind of model is assumed for the propagation of the wind field. The picture of a frozen turbulence, where the wind field forms a more or less homogeneous front and fluctuations transverse to the main flow direction are mostly neglected, is not appropriate for an accurate description on small time scales. In this regard, also the yawing dynamics, i.e. the process of adjusting the wind turbine's rotor with respect to the flow direction of the wind that is typically characterized by a finite response time, and a respective yawing misalignment may matter. A last point that shall be itemized here is the response of the used anemometers that similarly show a certain finite dynamical behaviour.

The objective of the standard procedure due to [IEC 2005a] is to average over all these dynamical effects and separate the actual characteristic large-scale behaviour of the WECS from the short-term fluctuations that are effected by the listed aspects of small-scale dynamics. But because of the asymmetry of the fluctuations that is introduced by the nonlinearity of the power curve if not present anyway, this kind of low-pass filtering defined by a simple averaging procedure is not an appropriate solution to the faced problem. I present a different approach in this work handling the asymmetries of the fluctuations as well as the dynamical influences on small scales much more flexible. For this purpose, I firstly elaborate on what may be called turbulent structures in wind speed and power output time series and how they can be characterized in section 3.2 before I introduce a phenomenological approach to the modelling of the effective dynamics in the power performance of a WECS in section 3.3.

3.2 Turbulent structures in wind speed and power output

3.2.1 Characterization of short-term structures in wind speed time series

Decomposition into a mean component and respective short-term fluctuations

A common approach in the study of turbulent flows is to decompose the considered variable into a mean value and corresponding short-term fluctuations around this value. The fluctuations that are obtained as difference between the measured time series and the derived average are often referred to as turbulence, and studying turbulence means, in this sense, investigating the properties of these fluctuations (cf. [Panofsky & Dutton 1984]). In principle, the definition of the averaging procedure is arbitrary. The most common choice is probably a time average. For the wind speed, for instance, we may write $u = \bar{u} + u'$ where \bar{u} is defined according to

$$\bar{u}_T := \frac{1}{T} \int_0^T dt u(t) \quad (3.8)$$

with regard to an averaging period of length T . For a time series of discrete data, the integration in (3.8) is replaced by a summation according to

$$\langle u \rangle_T := \frac{1}{N_T} \sum_{i=1}^{N_T} u(t_i), \quad (3.9)$$

where N_T is the number of discrete data points at times t_i that lie within the time period T . If not otherwise specified, \bar{u}_T and $\langle u \rangle_T$ are used synonymously here. To simplify matters, the index T is omitted in the following.

Not only the values for the steady wind speed \bar{u} but also the properties of the respective fluctuations u' essentially depend on the choice of the averaging period T . To illustrate this, figures 3.1(a) and (b) show an exemplary wind speed time series together with the averages \bar{u} and the fluctuations u' for two different values of T . Partly, the diagrams show considerable deviations which points out the difficulty that is related to the choice of an appropriate value for T . For a detailed description of the data that is used for these and the following figures see Appendix B.

For applications in the field of wind energy research, an averaging period of $T = 10$ min is commonly used (see e.g. [IEC 2005a] and section 3.1.2). This regulation is often justified with the existence of a spectral gap in the power spectrum of wind speed time series for time scales in this range. The observation of this so-called mesoscale gap goes back to [van der Hoven 1957] where it was localized as a region of low spectral density around periods of from approximately 5 min to 5 h, lying between the synoptic peak with high densities for large scales and the turbulent peak corresponding to a high spectral density for small scales. Over the last decades, the location of the mesoscale gap has been highly debated and its existence even totally denied. In [Panofsky & Dutton 1984] it is argued that even if such a minimum exists, the corresponding density will be far from zero which may cause severe difficulties in the problem of separating a mean from the turbulent flow. Accordingly, the application of such a separation is rather rigorous.

For the specific applications, as described above for the example of power performance testing according to [IEC 2005a], the actual data analysis is based on the average values whereas the short-term fluctuations are only considered in terms of the turbulence intensity. The turbulence intensity of a wind speed time series, defined by $I_u := \sigma/\bar{u}$ where σ is the standard deviation $\sigma := \sqrt{\langle [u' - \langle u' \rangle]^2 \rangle} = \sqrt{\langle u'^2 \rangle}$

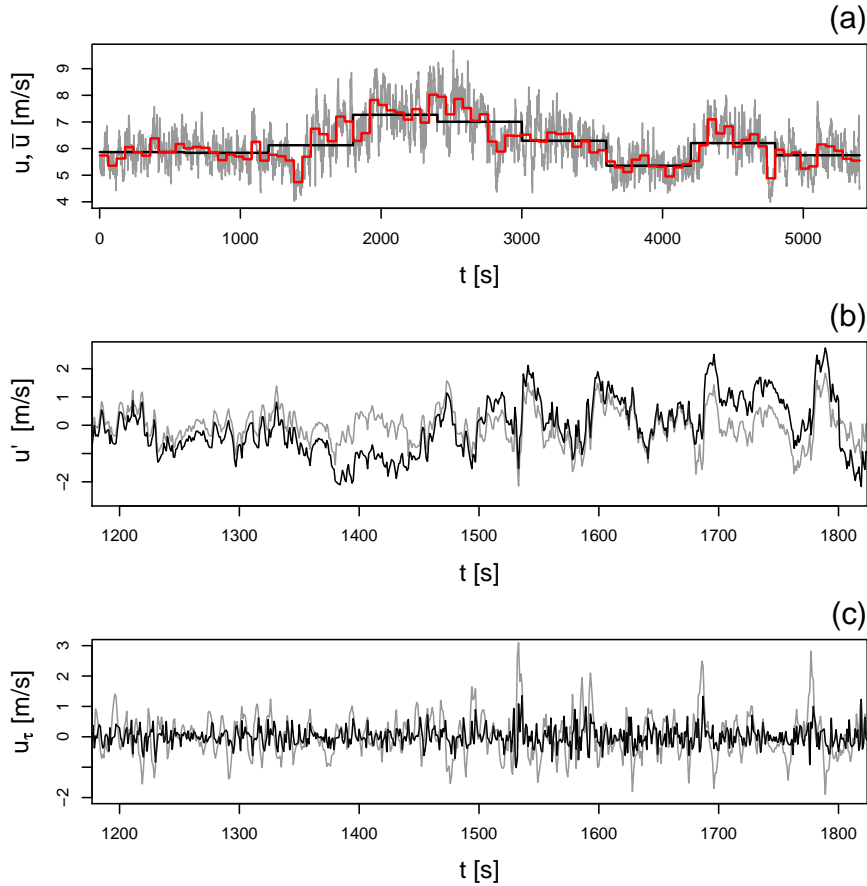


Figure 3.1: (a) Typical time series $u(t)$ of atmospheric wind speed (grey) together with the corresponding mean values \bar{u} for averaging times of $T = 10$ min (black) and $T = 1$ min (red) over a length of 90 min. (b) Resulting short-term fluctuations $u' = u - \bar{u}$ (black curve for $T = 10$ min and grey for $T = 1$ min) and (c) increments u_τ for values $\tau = 1$ s (black) and $\tau = 5$ s (grey), for a segment of 10 min in each case. (A detailed description of the used data is given in Appendix B.)

(note that $\langle u' \rangle$ vanishes due to its definition), is evaluated for each interval of length T . Its dependence on the mean value \bar{u} is utilized as a site specific characteristic (cf. the international standard IEC 61400-1 [IEC 2005b] specifying the design requirements for wind turbines, where this relation is used to define different wind turbine classes corresponding to respective categories of turbulence characteristics). A typical plot of I_u as function of \bar{u} is shown in figure 3.2(a).

Apart from the question if the utilized separation of mean and short-term dynamics is feasible, it should be checked how representative the turbulence intensity is for the investigated fluctuations. Its definition solely includes the first and second moment, i.e. mean value and standard deviation, of the considered data – a restriction that is only valid if the data is described by a normal or Gaussian distribution. To verify this assumption, I additionally determined the skewness S and the flatness (kurtosis) F of the single probability density functions (pdfs) that are defined as standardized third and fourth moment according to

$$S := \frac{\mu_3}{\sigma^3}, \quad F := \frac{\mu_4}{\sigma^4}, \quad (3.10)$$

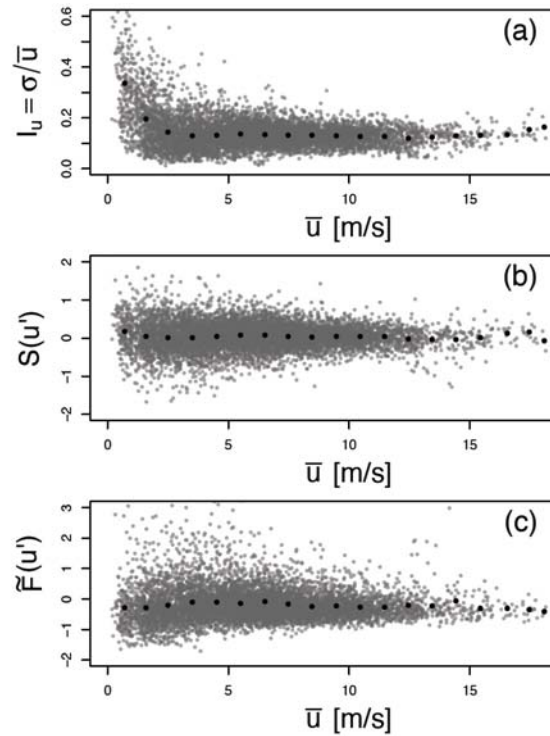


Figure 3.2: (a) Turbulence intensity I_u as fractional number (not in per cent) for a wind speed time series $u(t)$, consisting of 58 days of data sampled with a frequency of 1 Hz, as function of the respective mean wind speed \bar{u} . (b) Skewness S and (c) excess kurtosis \tilde{F} for the respective short-term fluctuations $u' = u - \bar{u}$ (or the high-frequency data u itself, resp.) conditioned on 10 min periods. The single values for each 10 min interval are given by the small grey dots. The black dots give bin averages for bins with a width of 1 m/s each.

where μ_n is the n -th centralized moment, i.e. here $\mu_n = \langle [u' - \langle u' \rangle]^n \rangle = \langle u'^n \rangle$, and σ again the standard deviation. Note that the second centralized moment equals the variance σ^2 , i.e. $\mu_2 \equiv \sigma^2$. A normal distribution has the skewness $S = 0$ and the kurtosis $F = 3$. Instead of F , often also the excess kurtosis $\tilde{F} := F - 3$ is introduced that equals zero for a normal distribution.

Figure 3.3 illustrates the results for three exemplary 10 min periods. The derived pdfs $p(u)$ conditioned on the particular time interval of length $T = 10$ min and respective mean value \bar{u} are compared to a normal distribution that is characterized by the standard deviation of the analyzed data. Calculated values for skewness and excess kurtosis are given in the caption. The corresponding results of the two measures for the complete data set are shown in figures 3.2(b) and (c). The values for both, the skewness and the excess kurtosis, lie approximately around zero which indicates that the high-frequency fluctuations for the single 10 min intervals are on average described by symmetric Gaussian shaped distributions. Though, the results in figures 3.3(a-c) show also considerable variations and respective deviations from these average levels. It is however questionable if these deviations from the normal distribution constitute the structures of non-Gaussian turbulence. The results in figures 3.2(a-c) and 3.3 primarily indicate that the number of data sampled over a period of only 10 min is too small to give well-defined, i.e. non-scattering, values for the three calculated measures.

The conclusion is different when the statistics of the fluctuations u' is not inves-

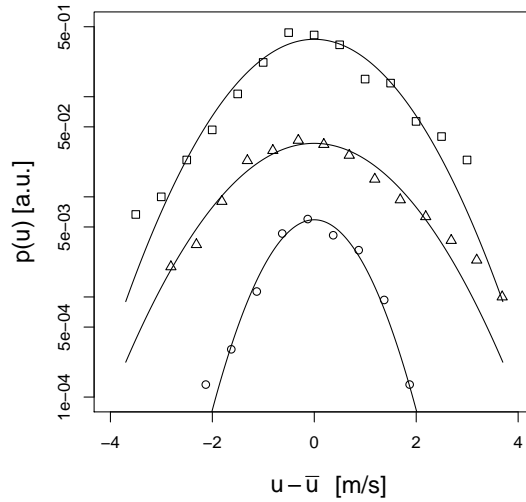


Figure 3.3: Pdfs of high-frequency data u conditioned on 10 min periods in comparison to normal distribution with same standard deviation – exemplarily for three different intervals and respective values \bar{u} . Results for turbulence intensity I_u , skewness S and excess kurtosis \tilde{F} are as follows: (\circ) $\bar{u}=5.88$ m/s, $I_u=0.12$, $S=0.01$, $\tilde{F}=0.01$; (\triangle) $\bar{u}=9.56$ m/s, $I_u=0.12$, $S=0.44$, $\tilde{F}=0.20$; (\square) $\bar{u}=8.75$ m/s, $I_u=0.12$, $S=0.17$, $\tilde{F}=0.54$. The two lower pdfs are shifted in vertical direction for clarity of presentation.

tigated with respect to the single 10 min intervals but analyzed combined as one data set. As shown in figure 3.4, the probability density function (pdf) of the entire set of fluctuations u' is characterized by a strongly non-Gaussian behaviour. The probability of large events is significantly higher than for a normal distribution with the same width, which is also referred to as heavy-tailed statistics or intermittency. For an averaging period of 10 min, the excess kurtosis of the fluctuations has a value of $\tilde{F} = 3.4$ that still exceeds the maximum values in figure 3.2(c) and is considerably higher than the average level of the values for the single 10 min intervals. Figure 3.4 also indicates how the entire distribution of fluctuations depends on the averaging period – the smaller the value of T , which defines the mean value \bar{u} with respect to that the fluctuations u' are determined, the more intermittent but at the same time narrower is the pdf of u' . The corresponding values for the standard deviation and the excess kurtosis, reflecting this behaviour, are given in the caption to figure 3.4.

The intermittency for the entire distributions of the fluctuations u' , in comparison to the distributions for the single intervals that show roughly Gaussian statistics, is traced back to the non-stationarity of the wind speed time series. The non-stationarity is a basic property of atmospheric wind speeds, and one of the main distinguishing features between atmospheric turbulence and the homogeneous, isotropic and stationary turbulence that is realized in laboratory experiments. To further illustrate this issue, I refer to the approach of describing turbulent structures in terms of increment statistics (cf. [Böttcher et al 2003, Böttcher et al 2007a]) that is presented in the following paragraph.

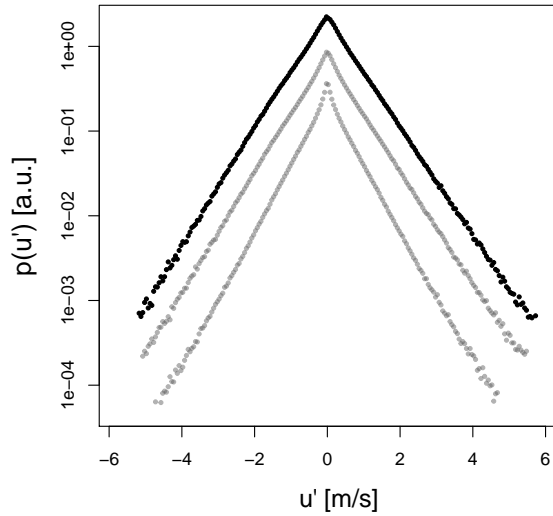


Figure 3.4: Pdfs of the fluctuations u' for the entire set of data (58 d, sampling rate 1 Hz) with respect to different values of T defining the averaging for u and consequently the determination of the fluctuations u' – from top to bottom $T = 10$ min (black), $T = 5$ min and $T = 1$ min (both in grey). The two lower pdfs are shifted in vertical direction for clarity of presentation. Values for the standard deviation and the excess kurtosis of the shown pdfs are $\sigma = 0.94$ m/s, 0.89 m/s, 0.74 m/s and $\tilde{F} = 3.4, 3.7, 4.5$ for $T = 10$ min, 5 min and 1 min (in this order).

Analyzing turbulent structures in terms of increments

A wind speed increment is defined according to

$$u_\tau(t) := u(t + \tau) - u(t) \quad (3.11)$$

as the difference between two values of wind speed that are separated by a time increment τ . An example of a time series $u_\tau(t)$ for two different values of τ is shown in figure 3.1(c). The difference between the short-term fluctuations u' and the increments u_τ can be seen as follows. While the fluctuations u' are per definition related to a reference value that is determined over a certain period of time but only indirectly to more than one distinct point in time, a time increment corresponds to a fluctuation that is determined by two distinct wind speed values and a respective time separation, and hence may reflect correlations with respect to the scale τ . Accordingly, increment statistics are also referred to as two-point statistics. They are often studied in terms of their moments $\langle u_\tau^n \rangle$ that are referred to as structure functions. It is straightforward to show that the second-order structure function is directly related to the auto-covariance of $\langle u(t + \tau)u(t) \rangle$ and respectively to the power spectral density, thus actually reflecting correlations in time. The skewness and kurtosis of the increment distributions are in turn defined as standardized third- and fourth-order structure functions.

The behaviour of increment pdfs for varying τ is quite similar to that of the pdfs of the fluctuations u' for different T . For increasing values of τ , the pdf of the increments u_τ becomes less intermittent while its width increases – see figure 3.5. This behaviour is described in more detail in figure 3.6, that shows the standard deviation as well as the logarithmic kurtosis of the increments u_τ as function of the time increment τ . For increasing values of τ the standard deviation increases

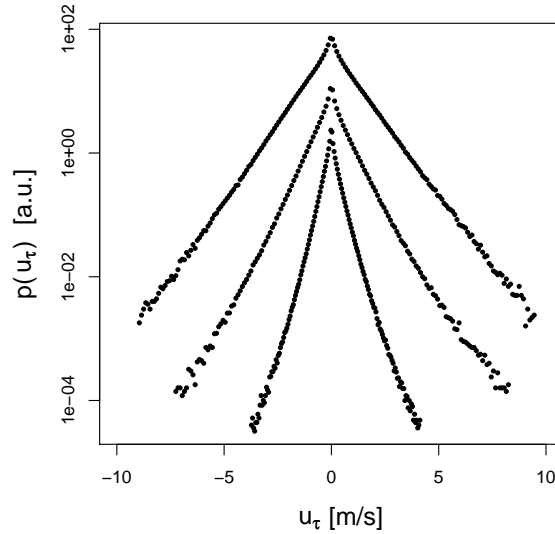


Figure 3.5: Pdfs of the wind speed increments u_τ for different time increments τ – from the bottom up $\tau = 1$ s, $\tau = 5$ s and $\tau = 30$ s. The two upper pdfs are shifted in vertical direction for clarity of presentation.

whereas the kurtosis decreases.

I plotted the function $\ln(F/3)$ instead of the actual kurtosis F for a specific reason that shall be briefly outlined in the following. As shown in [Beck 2004], the value $\ln(F/3)$ is proportional to the form parameter λ^2 that is introduced in [Castaing et al 1990] to fit the intermittent pdfs of wind velocity increments obtained from laboratory experiments. For this purpose, it is assumed that the intermittent pdf is given by a superposition of single Gaussian distributions with log-normal distributed variances according to

$$\begin{aligned} p(u_\tau) &= d\sigma p(u_\tau|\sigma)f(\sigma) \\ &= \int_0^\infty d\sigma \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{u_\tau^2}{2\sigma^2}\right] \times \frac{1}{\sigma\lambda\sqrt{2\pi}} \exp\left[-\frac{\ln^2(\sigma/\sigma_0)}{2\lambda^2}\right] \end{aligned} \quad (3.12)$$

where σ_0 is the median of the log-normal distribution and λ^2 its variance – all other quantities were defined before. The application of this formula as well as the utilization of $\lambda^2(\tau)$ as a characteristic quantity is well established in the field of turbulence research.

In [Böttcher et al 2003], this characterization is applied to reveal the basic difference between increment distributions for atmospheric and for laboratory turbulence data. For laboratory turbulence, the increment distributions show an intermittent behaviour for small values of τ but saturate against a normal distribution for large τ . For atmospheric data this saturation is not observed, at least not for the investigated time scales up to $\tau = 10^5$ s, and the values for λ^2 are overall larger. (Note that a larger value for λ^2 refers to a more intermittent distribution. A pdf with $\lambda^2 = 0$ has the shape of a Gaussian distribution.) This discrepancy partly vanishes when the wind speed increments are conditioned on a respective mean value. Thus, the non-stationarity of the atmospheric data is again of essential importance.

Due to this finding, Böttcher et al. proposed to model the intermittent pdfs of atmospheric wind speed increments as a superposition of different subsets of isotropic turbulence (see [Böttcher et al 2007a]). Denoting these subsets with $p(u_\tau|\bar{u})$ and

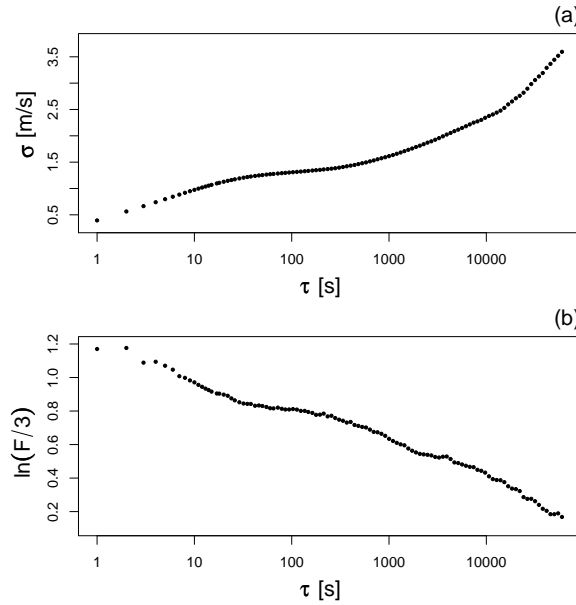


Figure 3.6: (a) Standard deviation σ and (b) form parameter $\lambda^2 \propto \ln(F/3)$ for the pdfs of the wind speed increments u_τ as function of the time increment τ (cf. figure 3.5).

the distribution of the mean values with $h(\bar{u})$, the pdf of the resulting increments is given by

$$p(u_\tau) = \int_0^\infty d\bar{u} h(\bar{u}) \times p(u_\tau|\bar{u}). \quad (3.13)$$

At this, $p(u_\tau|\bar{u})$ corresponds to the Castaing distribution (3.12), and $h(\bar{u})$ is assumed to be a Weibull distribution with certain shape and scale parameters that are derived from the corresponding set of experimental data. A respective decomposition of $p(u_\tau)$ into single conditioned pdfs $p(u_\tau|\bar{u})$ is shown in figure 3.7(a). The figure suggests that the width of $p(u_\tau)$ increases with the range of \bar{u} whereas the degree of intermittency decreases. The latter is even more distinct in figure 3.7(b), showing the pdfs normalized to their standard deviations. Note that the increments of atmospheric wind speed data are, thus, described in terms of two different kinds of superpositions – a superposition of single Gaussian distributions with log-normal distributed variances according to Castaing, and a second superposition of these so-called Castaing distributions with respect to different mean values of wind speed as a kind of non-stationary mixing. The two kinds of superpositions are, at the same time, related to two different sources of intermittency.

3.2.2 A note on gusts

The last preceding section has dealt with an appropriate characterization of wind speed fluctuations. Before I transfer the discussed concepts to the analysis of the respective power output data, I give a short note on their relevance for the definition of gusts.

In particular, I introduced two different concepts to describe short-term structures in turbulent wind speed time series – fluctuations with respect to a mean value defined by a certain averaging time T , and increment values determined by a time increment τ representing two-point statistics. Both approaches give a specific description of turbulence. In wind energy research, as well as in other rather appli-

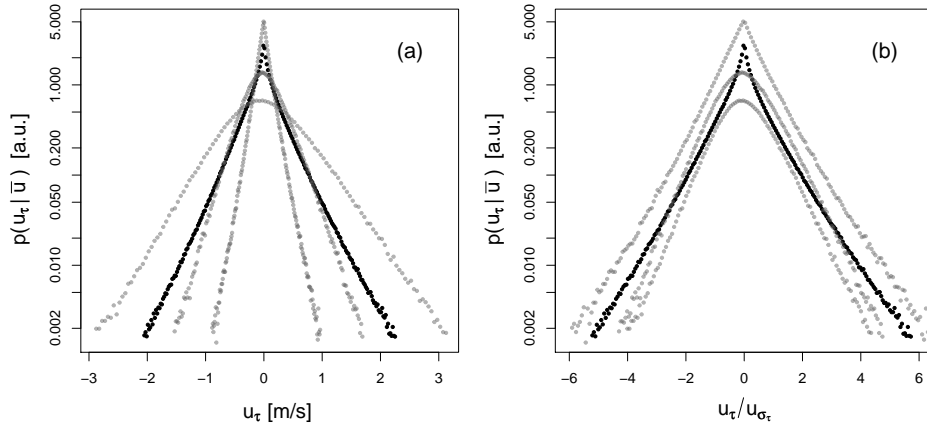


Figure 3.7: Entire pdf $p(u_\tau)$ (black) and the conditioned pdfs $p(u_\tau|\bar{u})$ for $\bar{u} \in [2, 4]$ m/s, $\bar{u} \in [5, 6]$ m/s and $\bar{u} \in [12, 25]$ m/s (from the center outwards, and all three in grey) for $\tau = 1$ s. Figure (b) shows the same pdfs as in (a) but with the increment values conditioned to their respective standard deviations.

cation oriented fields of research, turbulence and especially atmospheric turbulence is directly related to the occurrence of wind gusts (see e.g. [Risø-M-2632]). The frequency of occurrence, or more general the statistics of wind gusts, may define the degree of turbulence. Gusts are in turn nothing else than certain fluctuations, the explicit definition however is still an open issue and there exist a couple of different approaches.

In the standard IEC 61400-1 [IEC 2005b], a gust is defined as a temporary change in the wind speed, and it is proposed to characterize it by its rise-time, its magnitude and its duration. In the further details of the standard, several types of gusts are defined that are given by a precise description in terms of a theoretical function $u(t)$ for each case. The application of these definitions is rather difficult. In addition, they depend on at least two different reference periods or averaging times that considerably affect the results. To find the most appropriate averaging time is in general a non-trivial problem.

In [Böttcher et al 2007a], it is proposed to identify large increment values for small values of τ as wind gusts. This approach has the advantage that no averaging has to be defined and, furthermore, the corresponding increment statistics directly includes a statistics for extreme events, quantifying the probabilities determined by the so-called fat tails. The time increment τ corresponds to the rise-time of the gust. Increment pdfs for different values of τ thus define gust statistics for different rise-times. However, also the increment pdfs are not unambiguous. Similarly as the short-term fluctuations $u' = u - \bar{u}$, they depend on the sampling rate of the time series $u(t)$. A first averaging, for instance from 50 Hz to 1 Hz as for the data set I analyzed (see Appendix B), already modifies the data basis and influences the results for the pdfs of fluctuations.

By studying the pdfs of the wind speed increments as well as of the short-term fluctuations u' in the preceding section, I have shown that it is often not sufficient to describe the statistics of small-scale fluctuations or gusts only in terms of the second moment of the distributions. Rather, it may be necessary also to consider at least the fourth moment that defines intermittency and respectively the occurrence of large events. For the fluctuations u' , it has to be differentiated at this if respective statistics are investigated with respect to limited intervals in time as the typical 10

min periods or for a longer data set that is essentially affected by a non-stationarity of the considered quantity. A similar impact has the conditioning of the increments u_τ to a respective average wind speed value. To a large extent, intermittency is thus related to the non-stationarity of the wind speed time series – even though not solely.

Eventually, the question arises which kind of fluctuations actually act on a wind turbine that is exposed to a turbulent wind field – at this, a respective impact may be defined as a load or as well as an effect on the power output of the system. In general, gusts should be defined in an object-related way. For instance, the control system of the WECS may determine which kinds of fluctuations, and in particular with respect to which reference time, are relevant for the system. A similar argument applies for the power quality that is defined with respect to the grid the produced power is fed in. The definition of a specific gust directly depends on the kind of conclusion that shall be deduced from its value.

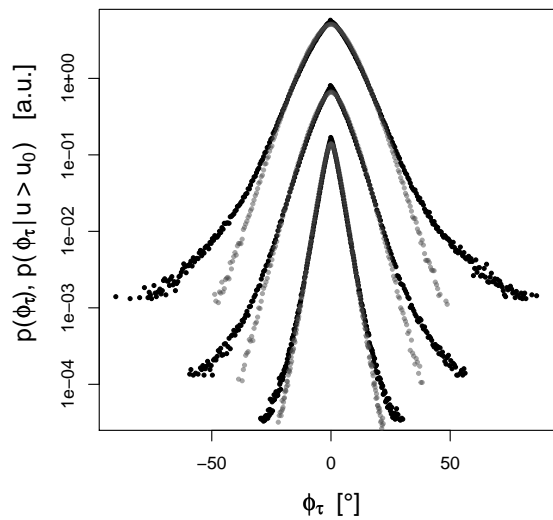


Figure 3.8: Pdfs of the wind directions increments ϕ_τ for all wind speeds (black) and for $u > u_0 = 4$ m/s (grey) – from the bottom up for $\tau = 1$ s, $\tau = 5$ s and $\tau = 30$ s. Values $|\phi_\tau| > 180$ are added or subtracted by 360° . The two lower pdfs are shifted in vertical direction for clarity of presentation.

Another aspect, that has not been discussed in this chapter yet, is the number of dimensions considered for the description of the wind field. The horizontal wind speed, that is solely studied so far, is a one-dimensional quantity, and thus gives a strongly reduced description of the three-dimensional wind field. To illustrate the magnitudes of the influences of the further wind velocity components, figures 3.8 and 3.9 show the statistics of fluctuations in the wind direction and the vertical wind velocity component that go along with the fluctuations in the horizontal wind speed discussed before. The increment values $\phi_\tau(t)$ and $u_{z,\tau}(t)$ are defined in accordance with (3.11). Their pdfs in figures 3.8 and 3.9 show a similar intermittent behaviour as those for the horizontal wind speed increments u_τ (cf. figure 3.6). Note that the broadened tails of the pdfs for the wind direction increments are mainly due to extreme direction changes at low wind speeds. Their physical impact as e.g. extreme load changes is therefore rather low. Figure 3.8 additionally shows the wind pdfs of the increments ϕ_τ for wind speeds $u > u_0$ ($u_0 = 4$ m/s), that are regarding their

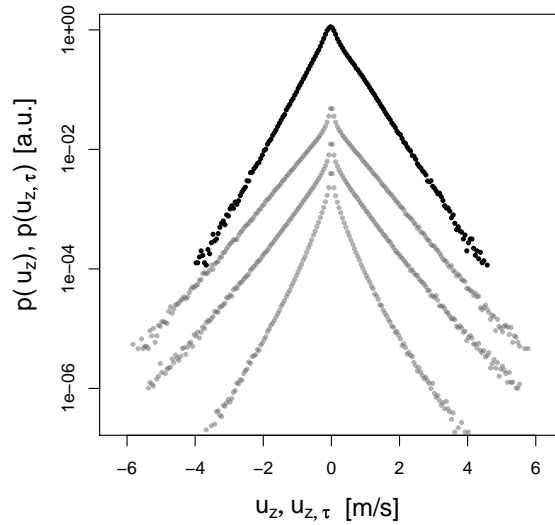


Figure 3.9: Pdfs of the vertical wind speed component u_z (black) and its increments $u_{z,\tau}$ for $\tau = 1$ s, $\tau = 5$ s and $\tau = 30$ s (from the bottom up, and all three in grey). The three lower pdfs are shifted in vertical direction for clarity of presentation.

shape much more similar to the pdfs of the wind speed increments shown above.

It becomes obvious that the picture of a wind field that forms a homogeneous front in horizontal direction is just as inappropriate as neglecting the fluctuations of the wind velocity in vertical direction. In the standard IEC 61400-1 [IEC 2005b], these three-dimensional fluctuations are accounted for by the specific gust definitions as that of an extreme direction change (EDC), an extreme wind shear (EWS), and an extreme coherent gust with direction change (ECD), that are introduced in addition to the extreme operating gust (EOG) describing only the extreme changes in the horizontal wind speed component. For all these gust types however apply the same critical remarks stated above.

3.2.3 Transfer of turbulent structures to the power output

In this section, I investigate to what extent the short-term fluctuations in the wind velocity are reflected in the power output of a wind turbine that is exposed to them. For this purpose, I again restrict my analysis to the horizontal wind speed as characterization of the wind field acting on the turbine, having in mind that this is only a reduced representation.

Just as the wind speed, the time series of the power output can be described in terms of a mean value and corresponding short-term fluctuations according to $P = \bar{P} + P'$ with an appropriate averaging period T . A turbulence intensity for the power output is accordingly defined by $I_P := \sigma/\bar{P}$ with $\sigma = \sqrt{\langle [P' - \langle P \rangle]^2 \rangle} = \sqrt{\langle P'^2 \rangle}$ as the standard deviation of the short-term power fluctuations. Applying an averaging period of $T = 10$ min, as proposed in the standard IEC 61400-12-1 for power performance testing, I again derived the turbulence intensity as well as the skewness and the excess kurtosis for the distribution of the fluctuations P' for the single 10 min intervals of a measured set of power data. (The data was measured in parallel to the wind speed data analyzed in 3.2.1 – for a detailed description see Appendix B.) The results are shown in figures 3.10(a–c) for the whole data

set and in figure 3.11 for three exemplary intervals of a length of 10 min each – analogously to figures 3.2 and 3.3 for the wind speed u . In all three measures, turbulence intensity, skewness and kurtosis, the influence of the power curve as transfer function between wind speed and power output fluctuations is clearly seen. The values for the turbulence intensity of the power output are overall larger than those for the wind speed but significantly decrease in the range of rated power – see figure 3.10(a). Similarly, as indicated in figures 3.10(b) and (c), the skewness and the kurtosis show a specific behaviour in this range, which characterizes the transition region from full load to partial load at the point of rated wind speed as a very essential region. In particular, the systematic deviations of the skewness from zero indicate the asymmetries of power fluctuations that are caused by the nonlinear parts of the power curve. These asymmetries give rise to the systematic errors in the reconstruction of power curves applying the standard procedure due to IEC 61400-12-1, that are discussed in 3.1.3. At the same time, figure 3.11 indicates that the pdfs of the power data within the single 10 min intervals are far from well-defined by the derived measures.

I have further investigated the small-scale behaviour of the power output by studying the increments

$$P_\tau(t) := P(t + \tau) - P(t) , \quad (3.14)$$

where τ again defines the respective value of time separation for the considered two-point statistics. Figure 3.12 shows the increment pdfs for three different values of τ . Comparing the pdfs for the power increments with those for the wind speed in

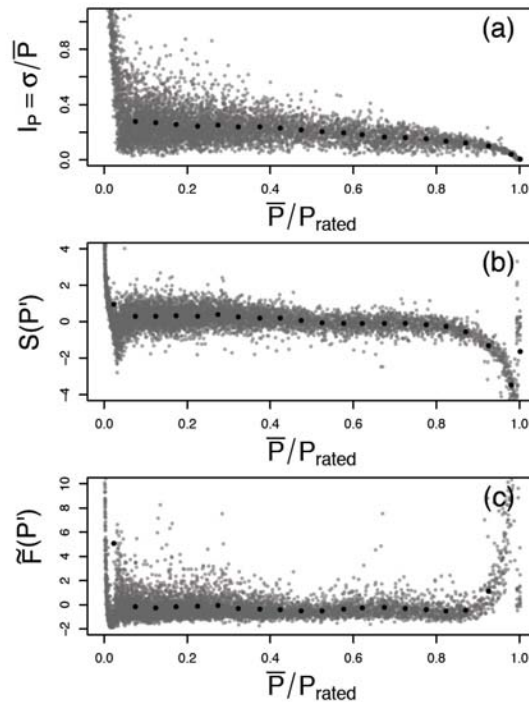


Figure 3.10: (a) Turbulence intensity I_P for a power output time series $P(t)$, consisting of 58 days of data sampled with a frequency of 1 Hz (for a detailed description see Appendix B). (b) Skewness S and (c) (excess) kurtosis \tilde{F} for the respective short-term fluctuations $P' = P - \bar{P}$. The single values for each 10 min interval are given by the small grey dots. The black dots give bin averages for bins with a width of $0.05P_{\text{rated}}$.

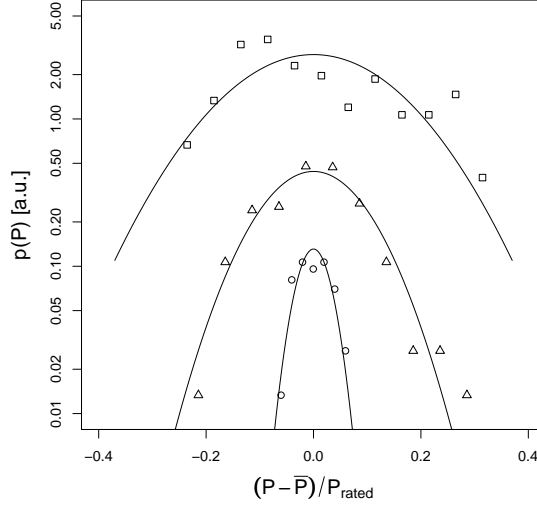


Figure 3.11: Pdfs of high-frequency data P conditioned on 10 min periods in comparison to normal distribution with same standard deviation – exemplarily for three different intervals and respective values \bar{P} . (Examples for the same intervals as in figure 3.3, symbols are taken over.) Results for turbulence intensity I_P , skewness S and excess kurtosis \tilde{F} are as follows: (○) $\bar{P}/P_{\text{rated}}=0.17$, $I_P=0.18$, $S=0.07$, $\tilde{F}=-0.91$; (△) $\bar{P}/P_{\text{rated}}=0.69$, $I_P=0.13$, $S=0.24$, $\tilde{F}=0.29$; (□) $\bar{P}/P_{\text{rated}}=0.46$, $I_P=0.32$, $S=0.51$, $\tilde{F}=-0.79$. The two lower pdfs are shifted in vertical direction for clarity of presentation.

figure 3.5, the additional structures for $p(P_\tau)$ with $\tau = 1$ s and $\tau = 5$ s, denoted in the following as bumps, attract attention. I studied these bumps, corresponding to power jumps of approximately a third and a fifth of the rated power, in more detail and found that they occur for a wide range of power values $P(t)$ but most frequently as a jump from or to rated power. They vanish at about $\tau = 20$ s – for this reason we have related them to a discretization effect (cf. [Gottschall & Peinke 2007]). Note that these structures are specific for the investigated WECS and that the pdfs of power increments for other types of machines may look quite different.

Apart from these additional structures, the pdfs for the power increments show a similar behaviour and especially a similarly distinctive intermittency as the increment pdfs for the wind speed, and I conclude that the turbulent small-scale structures in the wind speed are more or less directly transferred to the power output of the considered wind turbine.

To inspect this transfer in more detail and roughly quantify the extent to which the WECS integrates over the wind speed fluctuations, I pursued the following simple approach. I compared the measured power increments with the reconstructed increments of a power output time series that results from simply applying a power curve derived according to IEC 61400-12-1 to the measured wind speed time series – i.e. $P_{\text{p.c.}}(t) := P_{\text{p.c.}}(u(t))$ where $P_{\text{p.c.}}(u)$ is the measured and splined or interpolated power curve. As shown in figure 3.13 for a time separation of $\tau = 20$ s, the pdf for the reconstructed power increments $P_{\text{p.c.,}\tau}$ (defined analogously to (3.14)) is considerably wider than that for the measured increments P_τ . This deviation is expressed in more detail in figures 3.14(a) and (b), showing the standard deviation and the logarithmic kurtosis $\ln(F/3)$ of the measured and reconstructed increment pdfs as a function of the time separation τ . The pdfs for the measured power increments are over a wide range of values τ narrower, i.e. characterized by a smaller standard deviation, but at the same time more intermittent due to a larger kurtosis

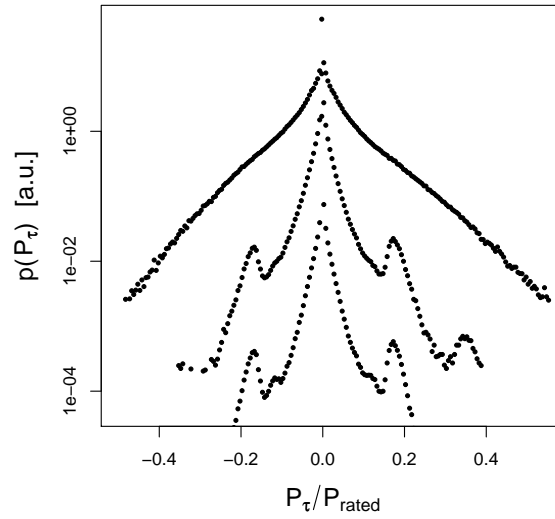


Figure 3.12: Pdfs of the increments P_τ for different time increments τ – from the bottom up $\tau = 1$ s, $\tau = 5$ s and $\tau = 30$ s. The two lower pdfs are shifted in vertical direction for clarity of presentation.

than the pdfs for the reconstructed increments. Note that the large values of $\ln(F/3)$ for the measured increments in the range of τ from 1 s to approximately 20 s are caused by the additional structures denoted as bumps above.

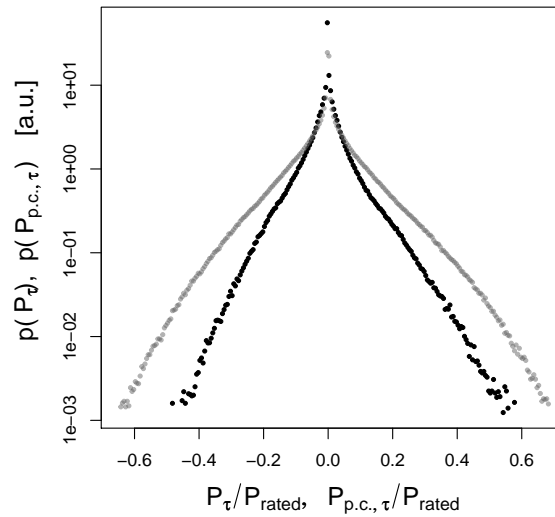


Figure 3.13: Pdfs for the measured and reconstructed power increments, P_τ (black) and $P_{p.c.,\tau}$ (grey), respectively, for $\tau = 20$ s.

The simple reconstruction model I applied is determined as a model without any relaxation of power output, i.e. an instantaneous response of the power to the wind speed. Consequently, I have concluded that the observed deviations between the model and respective measurements are due to a finite relaxation effect that is certainly inherent in the measured data. It is straightforward to show that a

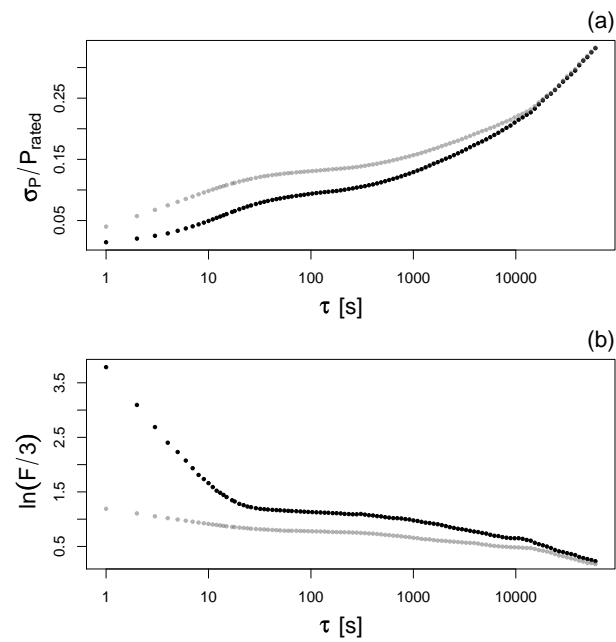


Figure 3.14: (a) Standard deviation σ_P and (b) logarithmic kurtosis $\ln(F/3)$ for the pdfs of the increments P_τ (black) and $P_{p.c.,\tau}$ (grey) as function of the time increment.

simple relaxation model for the power response introduces increment pdfs that are narrower but more intermittent than those for simply applying a power curve on the wind speed time series. I propose such a relaxation model in the next section, and show how it can be utilized to define a dynamical power characteristic.

3.3 Definition and reconstruction of a dynamical power characteristic

3.3.1 Stochastic relaxation model

According to [Anahua et al 2008], I propose to model the dynamics of a wind turbine's power output as a set of univariate Langevin processes given by the stochastic differential equation

$$\frac{d}{dt}P(t) = h_i(P, t) + \sqrt{g_i(P, t)} \Gamma_i(t) \quad (3.15)$$

for each fixed value of wind speed u_i . The terms h_i and g_i are arbitrary functions of the power output P at time t , and $\Gamma_i(t)$ is a noise process. The wind speed is a parameter, as indicated by the index i , but does not act as a second variable, so that this model corresponds to a quasi-onedimensional approach. The conditioning on the wind speed is realized by applying a wind speed binning with a fixed bin width, as it is introduced in the standard IEC 61400-12-1 [IEC 2005a]. The deterministic part of the dynamical equation (3.15) describes the relaxation behaviour of the system with respect to one or more steady states that are also referred to as fixed points. For only one fixed point and a linear relaxation, we write

$$h_i(P, t) = -\alpha_i[P(t) - P_{\text{FP}}(u_i)] \quad (3.16)$$

where α_i is a relaxation constant and $P_{\text{FP}}(u_i)$ denotes the fixed point in the wind speed bin around u_i . For a more general relaxation behaviour, α_i may refer to an arbitrary nonlinear function of the difference between the actual power value and the respective steady state. The wind speed fluctuations act as a noisy driving force in this model, and induce a kind of switch-over between the single wind speed bins and the respective one-dimensional Langevin models.

Due to the assumption that (3.15) describes a Langevin process and $\Gamma_i(t)$ accordingly corresponds to a Langevin force defined as Gaussian distributed and δ -correlated white noise (with $\langle \Gamma_i(t) \rangle = 0$ and $\langle \Gamma_i(t_1) \Gamma_j(t_2) \rangle = 2\delta_{ij}\delta(t_1 - t_2)$), the functions $h_i(P, t)$ and $g_i(P, t)$ equate drift and diffusion coefficients, i.e. $h_i(P, t) \equiv D_i^{(1)}(P, t)$ and $g_i(P, t) \equiv D_i^{(2)}(P, t)$, and can be reconstructed directly from a corresponding set of data – as discussed in Chapter 1 (section 1.2). I elaborate on the application of this reconstruction scheme in the next section. Firstly, I want to discuss how this simple model can be utilized to study typical effects that are found in the context of estimating power curves for wind turbines.

In [Böttcher et al 2007b], the relaxation model (3.15, 3.16) is integrated together with a theoretical power curve that is defined according to

$$P_{\text{FP}}(u) := \begin{cases} au^3 & \text{for } u < u_{\text{rated}} \\ P_{\text{rated}} & \text{for } u \geq u_{\text{rated}} \end{cases} \quad (3.17)$$

to obtain numerical power output data. The constant a is related to the values of rated power and rated wind speed according to $a = P_{\text{rated}}/u_{\text{rated}}^3$, determining a continuous transition. The wind speed data is either simulated as well, or a measured time series is used. Figure 3.15 illustrates the considered model. The arrows indicate the drift field, i.e. the values $D_i^{(1)}(P_j)$ for each wind speed bin u_i and each power bin P_j , describing the relaxation towards the theoretical power curve. The power curve corresponds with the discrete fixed points $P_{\text{FP}}(u_i)$ for the single wind speed bins.

I have applied this model to demonstrate the systematic errors of a power curve that is reconstructed by simple averaging procedures, and their relation to the

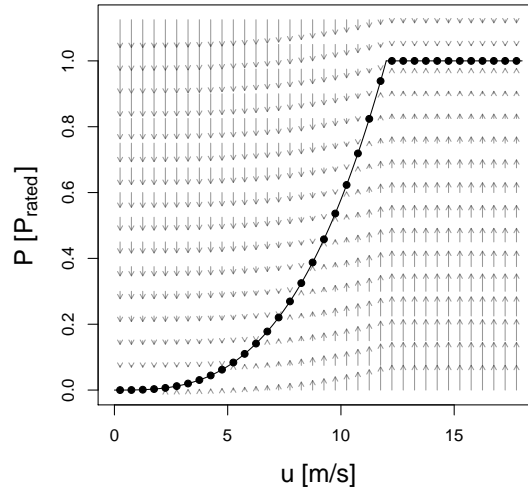


Figure 3.15: Illustration of relaxation model. The arrows indicate the values of the drift field $D_i^{(1)}(P_j)$ for the wind speed bin u_i and the power bin P_j , that describes the relaxation towards the theoretical power curve (given by the line) – here for a linear relaxation function with $\alpha = 0.1$. The dots denote the fixed points $P_{FP}(u_i)$ for the single wind speed bins.

relaxation effect. For this purpose, I took a time series of measured wind speed data (10 successive days, sampled with a frequency of 10 Hz and with an average turbulence intensity of $I_u \approx 0.13$ – cf. Appendix B) and simulated a respective time series of power output. I considered a linear relaxation model according to (3.15) and (3.16) with $\alpha_i = \alpha = 0.1$ and $g_i(P, t) = \beta = 0.01$ for each wind speed bin u_i , and the theoretical power curve defined in (3.17) with $u_{rated} = 12$ m/s and $P_{rated} = 1$. From the simulated data, I firstly reconstructed a power curve due the IEC 61400-12-1 standard [IEC 2005a], i.e. calculating the average values over periods of 10 min for the wind speed and power output and subsequently applying the method of bins. As second approach, I calculated a short-term averaged power curve by directly applying the method of bins to the high-frequency data. These two curves are shown in figures 3.16(a–c) as dashed and dot-dashed lines in comparison to the theoretical power curve (defined according to (3.17)) that is given by the solid line. Both reconstructed curves show significant deviations from the theoretical curve, especially in the nonlinear regions of the power curve. I simulated a second time series of power output by simply inserting the wind speed data into the definition of the theoretical power curve, and again derived the power curve due to IEC 61400-12-1 as well as the short-term averaged curve from the simulated data set. The results are shown as open circles and triangles in figures 3.16(a–c). The corresponding power curve due to the IEC standard for this second data set shows almost exactly the same deviations as that for the first data set, whereas the deviations for the short-term averaged curve vanish.

This simple example reveals that the systematic errors of the estimating procedure due to IEC 61400-12-1 are definitely caused by the averaging over periods of 10 min and neglecting at this the asymmetric transfer of fluctuations from the wind speed to the power output. The deviations of a power curve that is directly reconstructed from high-frequency data however result from the finite response behaviour of the system. The latter vanish when the sampling rate of the data, to that the

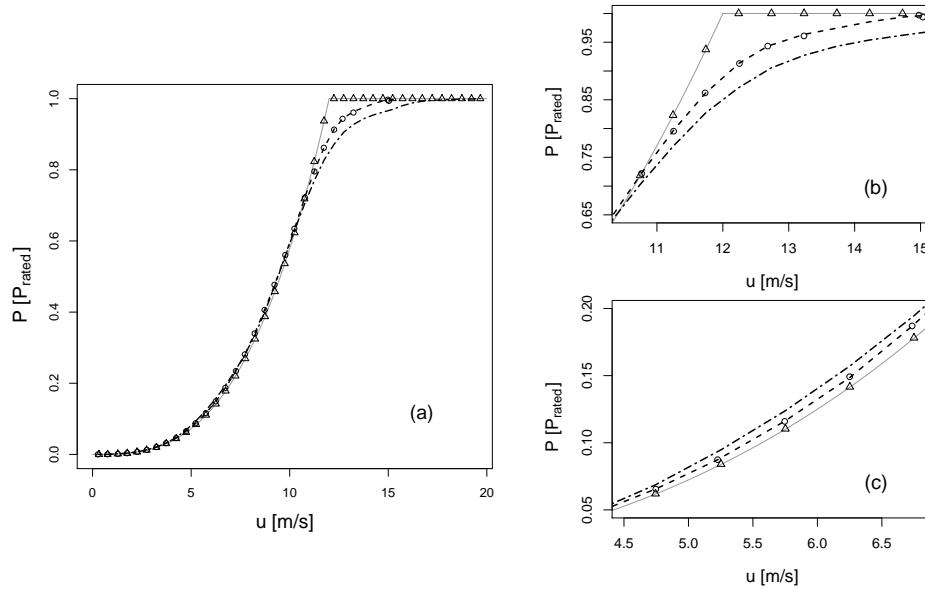


Figure 3.16: Reconstructed power curves for measured wind speed and numerical power output data with (broken lines) and without relaxation (symbols) – for details about the two used models see the text – in comparison to the theoretical curve (solid line, grey). Dashed line and open circles for reconstruction according to the IEC 61400-12-1 standard, dot-dashed line and open triangles for applying the method of bins directly to the high-frequency data. Figures (b) and (c) are close-up views of the entire curves in (a).

method of bins is applied, is larger than the time constant of the relaxation dynamics – with the drawback that then the above mentioned averaging errors emerge.

In [Burton et al 2001], the observed effect of rotating the power curve in clockwise direction, found here as consequence of the finite relaxation dynamics in the high-frequency data, is illustrated by considering the correlation between wind speed and power output data for high sampling rates. A respective correction is proposed that accounts for the respective correlation coefficient that is derived on the basis of the measured data. That is, the final effect of the decorrelation is utilized to find an appropriate correction but not the actual cause in terms of the relaxation dynamics. In this thesis, in contrast, a reconstruction scheme is presented that is based on the latter aspect.

3.3.2 Reconstruction of dynamical power characteristic

If the power conversion dynamics of a wind turbine, i.e. the process described by the time series $P(t)$ in relation to a wind speed $u(t)$, is actually given by a set of Langevin processes (3.15), it is straightforward to apply the reconstruction scheme outlined in Chapter 1 of this thesis, and reconstruct the effective dynamical equations directly from a set of measured data. In particular, the obtained relaxation dynamics determined by the drift coefficient for each wind speed bin and the respective fixed points can be utilized to define a power characteristic for the investigated WECS. In this regard, the definition of a dynamical power curve was introduced in [Anahua et al 2008]. In this section, I outline this approach with some minor modifications and illustrate its application for a set of measured data. The method is basically organized in five steps – the data selection, the calculation of the conditional moments, the estimation of the drift coefficients, the fixed-point analysis,

and finally the presentation of the results. The data we analyzed consists of wind speed, wind direction and power output time series, measured simultaneously over a period of 22 successive days and sampled with a frequency of 1 Hz – for a detailed description see Appendix B. A discussion of the model, especially with regard to its applicability, possible restrictions and specific requirements from the theoretical background, is given in the following section.

Data selection Requirements regarding the quality of the data are adopted from the standard IEC 61400-12-1 [IEC 2005a], and may be considered in terms of respective category B uncertainties if the necessary information is available. For the data set analyzed here, this kind of information has been not available, so that the corresponding part of evaluation of uncertainty had to be neglected. The effective sampling rate of the data shall be in the range of seconds, a minimum frequency of 1 Hz is suggested from our side. In general – to apply the method of reconstruction, it is necessary that the considered process is oversampled with respect to its relaxation time. Based on experience, a factor of 10 is at this a minimum requirement. Preferably, the data shall be sampled successively, discontinuities must be marked in any case. Data that is considered to be not valid due to a corresponding status signal for the WECS or similar criteria is rejected. In the same manner, data points are excluded for that the measured wind direction lies outside the valid wind sector that is defined in accordance with IEC 61400-12-1. The data selection is directly implemented in the algorithm. If air density or pressure and temperature data is available, a data normalization either for the wind speed or the power data is recommended. The remaining data is binned for the analysis with respect to both the wind speed and the power values. For the wind speed binning a bin width of 0.5 m/s is fixed, the width of the power bins is variable. The data shall be evenly distributed over the range of wind speed values. It is recommended to illustrate the distribution of valid data e.g. by means of a density plot.

Calculation of conditional moments For each wind speed and each power bin the conditional moments

$$M_{ij}^{(n)}(P, \tau) = \langle [P(t + \tau) - P(t)]^n | P(t) = P_j, u(t) = u_i \rangle \quad (3.18)$$

($n = 1, 2$) are calculated for different values of τ . For the binning only the points $u(t)$ and $P(t)$ are decisive but not the values at time $t + \tau$. The errors for the conditional moments are defined as common standard errors.

Estimation of drift coefficients To estimate the coefficients $D_i^{(1)}(P)$ and $D_i^{(2)}(P)$, the values for the conditional moments are fitted by a linear regression over a specified range of values $\tau \in [\tau_1, \tau_N]$, as illustrated in figure 3.17. The slope of the linear fit gives an estimate for the respective coefficient. Note that this is an approximation of the limit

$$D_i^{(n)}(P) = \frac{1}{n!} \lim_{\tau \rightarrow 0} \frac{1}{\tau} M_{ij}^{(n)}(P, \tau) \quad (3.19)$$

(cf. Chapter 2). The range for the fitting is determined as relevant range of time scales for the considered dynamics, but its definition is to some extent arbitrary. It constitutes a characteristic parameter for a specific type of WECS and a particular measurement setup.

To the reconstruction of the power characteristic only the values of the drift coefficients $D_i^{(1)}(P)$ are of interest, the values of $D_i^{(2)}(P)$ however are used to determine

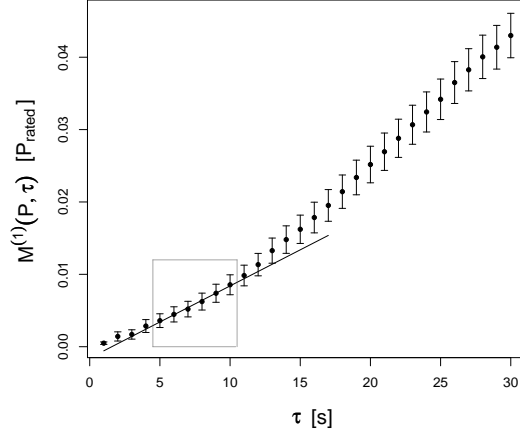


Figure 3.17: Estimation of the drift coefficient by applying a linear fit to the first conditional moment for a specified range of τ . Here for the wind speed bin $u_i=(6.75\pm 0.25)\text{m/s}$ and the power bin $P_j/P_{\text{rated}}=0.05\pm 0.01$ of the analyzed set of measured data. As range for the fitting, we determined $\tau \in [5 \text{ s}, 10 \text{ s}]$ (marked by the grey box).

the errors $\sigma[D_i^{(1)}(P)]$. According to [Kleinhans & Friedrich 2007], the standard error of the drift coefficient for a Langevin process is given by

$$\bar{\sigma}[D_i^{(1)}(P)] = \sqrt{\frac{2}{\tau} \frac{D_i^{(2)}(P)}{N} - \frac{(D_i^{(1)}(P))^2}{N}}, \quad (3.20)$$

where N is the number of data points in the bin. The first term corresponds to a common standard error for a set of independent random numbers. The reduction by the second term stems from the finite correlations induced by the drift term. To obtain the total error for the estimated drift coefficient, we have to additionally consider the error of the linear regression. The entire error $\sigma[D_i^{(1)}(P)]$ results from a squared summation of the two uncertainties.

Fixed-point analysis The fixed points of the relaxation dynamics are defined as points with zero drift, i.e.

$$D_i^{(1)}(P_{\text{FP}}) \equiv 0. \quad (3.21)$$

For the present analysis, only the stable fixed points determined by a negative slope $dD_i^{(1)}(P_{\text{FP}})/dt < 0$ are considered. Figure 3.18(a) shows the behaviour of the drift coefficient as function of the power values for an exemplary wind speed bin. The discretization of the horizontal axis due to the power binning as well as the errors in vertical direction make a precise estimation of the fixed point as zero-crossing of the drift function difficult. To overcome this problem, Anahua et al. proposed to derive instead the drift potential

$$\phi_{D_i^{(1)}}(P) := - \int^P dP' D_i^{(1)}(P'), \quad (3.22)$$

apply either before or after the integration a spline, i.e. a polynomial interpolation, and determine the fixed points as the minima of $\phi_{D_i^{(1)}}(P)$ [Anahua et al 2008]. The problem of this procedure is however that the lower limit of the integral is not explicitly defined but may be chosen arbitrarily. In the same manner, the absolute values for the uncertainty of the drift potential depend on the definition of the lower

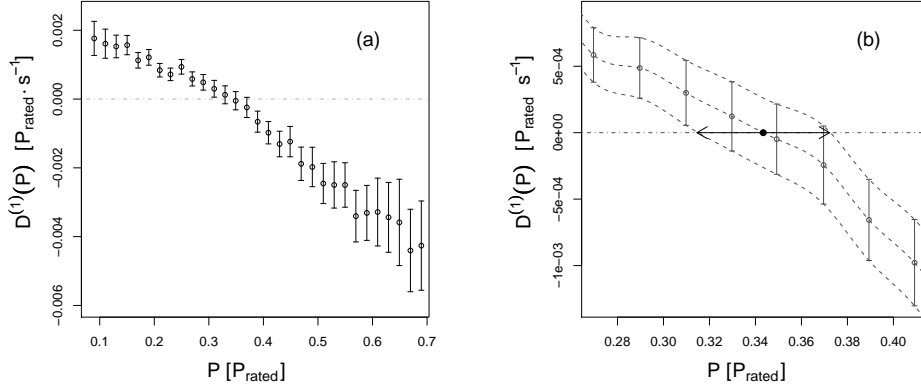


Figure 3.18: Illustration of fixed-point analysis. Figure (a) shows the values of the drift coefficient $D_i^{(1)}(P_j)$ for the single power bins P_i in the wind speed bin $u_i=(7.25\pm 0.25)\text{m/s}$. Figure (b) is a close-up view and exemplifies the estimations of the fixed point by applying cubic splines to the values $D_i^{(1)}(P_j)$ as well as to its upper and lower limits (dashed lines). The resulting fixed point is given by the black dot on the horizontal line of zero drift, the respective error by the arrow.

bound for the integral. On this account, I decided to estimate the fixed points not from the drift potential but directly from the drift coefficient after applying a highly resolved spline. This procedure of a local interpolation is illustrated in figure 3.18(b). The values for the drift coefficient as well as the upper and the lower limits of the respective errors are interpolated by cubic splines. For each of the three resulting curves the point with the smallest deviation to zero is determined, corresponding to the fixed point and its upper and lower limit.

Presentation of the results The dynamical power characteristic is given by the fixed points, that also represent the dynamical power curve, together with the drift field that consists of the values $D_i^{(1)}(P_j)$ for each wind speed bin u_i and each power bin P_j . Figures 3.19(a) and (b) show the respective results for the analyzed data set for two different power binnings. It is obvious that the denoted errors for the fixed points are essential for the evaluation of the results. They determine, for instance, if an obtained fixed point corresponds to a clearly defined steady state or rather a less distinct band. Similarly, particular fixed points may be evaluated by considering them in respect to the displayed drift field that reflects the distribution of data. In this way, boundary effects are potentially revealed.

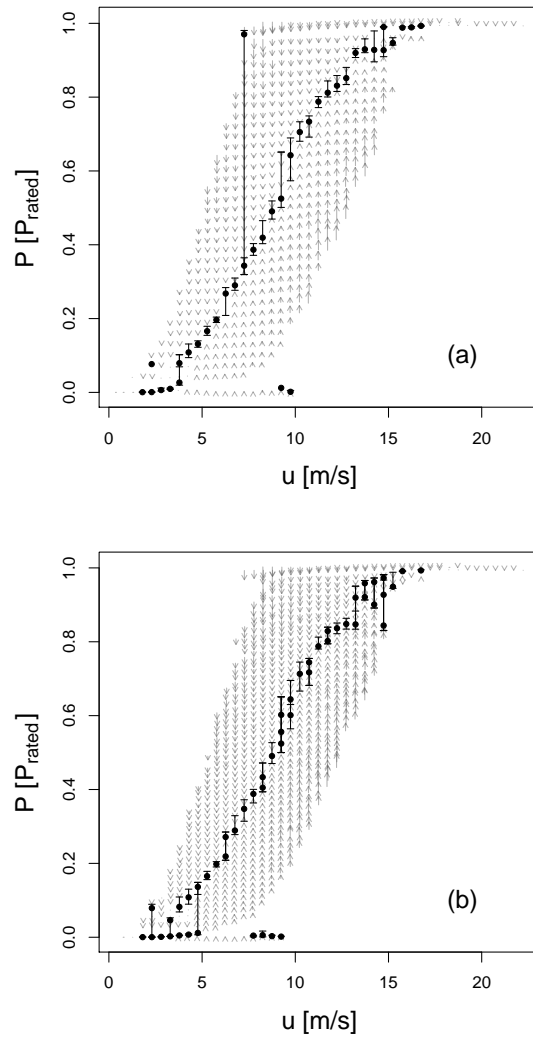


Figure 3.19: Dynamical power characteristic for two different power binnings – (a) with a bin width of $P_{\text{rated}}/35$, and (b) $P_{\text{rated}}/60$. The arrows denote the drift field defined by the values $D_i^{(1)}(P_j)$ for each wind speed bin u_i and each power bin P_j , and the dots give the (stable) fixed points of the relaxation dynamics.

3.4 Discussion of the model

The dynamical model for the power performance of a WECS, that is proposed in this chapter, corresponds to a phenomenological approach, i.e. the system is solely described on the basis of the empirical data without making use of further theoretical assumptions than the general stochastic model. The WECS is considered to be a complex system, and its power performance dynamics is modelled in an effective way but not in all details. The separation of the dynamics into a deterministic and a stochastic part, as given by the Langevin approach, allows for an at least rough classification of system-specific and site-dependent effects – where the former are given by the drift and the latter by means of the diffusion in the proposed model. Incorporating a stochastic component in the dynamical description, in this way, enables a flexible modelling of the impact of the turbulent wind structures on the power performance.

At the same time, it is obvious that the proposed model has certain limits that, correspondingly, may affect the robustness and the reliability of the obtained dynamical power characteristic. It is the purpose of this section to discuss these limits and provide in some respects a scope for the applied modelling. At this, I dwell on the quality of the results by presenting three different self-consistency tests for the obtained results, the applicability of the model with regard to technical as well as physical aspects discussing in particular the determination of an appropriate time scale of description as well as the representativity of the considered wind speed, and finally comment on the benefit of the dynamical approach, in particular with respect to the relevance of the reconstructed dynamical power characteristic.

3.4.1 Self-consistency

With self-consistency the accuracy of the reconstructed results in comparison to the intrinsic behaviour that actually determines the investigated process is meant. There are different ways to test the self-consistency of the reconstructed effective dynamical model and the results for the dynamical power characteristic – I present three different approaches here.

The first approach is to test the method of reconstruction on the basis of numerical data obtained from the introduced relaxation model (3.15, 3.16) and the theoretical power curve defined according to (3.17). Results for the reconstructed fixed points in comparison to the theoretical curve used for the simulation are shown in figures 3.20(a–c). For the most regions of the power curve, the reconstructed values well agree with the theoretical ones. In the range of the transition from partial load to full load at rated wind speed, i.e. at the strongest nonlinearity of the power curve, there are however significant deviations. These deviations are due to the non-stationarity of the wind speed in combination with the discrete wind speed bins. Accordingly, they increase with the turbulence intensity of the wind speed data. They are particularly pronounced for a power curve with a sharp non-differentiable edge as in the definition (3.17). For a smoother transition, that is probably more realistic for experimental data, the deviations are considerably smaller and may be neglected in most instances (see [Gottschall & Peinke 2007], as well as [Gottschall & Peinke 2008a] and Appendix A).

For measured data, a comparison of the reconstructed fixed points with a theoretical power curve is normally not possible. Even the power curves that are provided by the manufacturers are, in general, only either measured curves or based on simple blade element momentum (BEM) models. What can be done for our stochastic approach, and this is the second procedure of testing the self-consistency of the obtained results I present here, is to check the assumptions for the application of the Langevin model with the results from the reconstruction. That is,

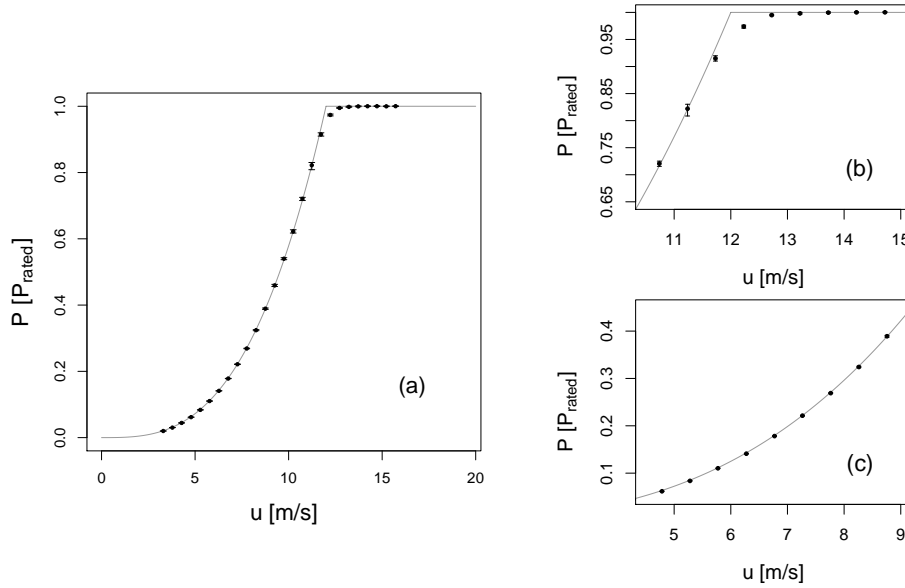


Figure 3.20: Reconstructed fixed points (symbols with error bars) for numerical power output data, simulated on the basis of measured wind speed data (10 successive days, sampled with a frequency of 10 Hz and with an average turbulence intensity of $I_u \approx 0.13$) and a theoretical power curve (solid line) as input for the simulation – see (3.15) with (3.17). Figures (b) and (c) are close-up views of the entire curve in (a).

one reconstructs the noise process $\Gamma_i(t)$ for a fixed wind speed bin u_i by transforming (3.15) and inserting the derived results for the drift and diffusion coefficient, $D_i^{(1)}(P_j)$ and $D_i^{(2)}(P_j)$, respectively. Subsequently, one investigates its properties, in particular the respective probability density and autocorrelation functions. For the application of the Langevin approach, a δ -correlated and Gaussian distributed white noise process is required. For the data I analyzed in 3.3.2 (see also Appendix B), I reconstructed $\Gamma_i(t)$ for different wind speed bins, each with a bin width of 0.5 m/s. The requirement of the δ -correlation is approximately fulfilled for each investigated case, deviations in terms of a non-vanishing correlation coefficient for non-zero time differences between two signals are within the uncertainty of measurement. However, the assumption of Gaussian pdfs for the single time series $\Gamma_i(t)$ does not apply, as exemplarily shown in figure 3.21 for three different wind speed bins u_i . In addition to an intermittent shape, the pdfs are characterized by additional peak values around zero and non-vanishing probabilities for large values of Γ_i . The distinctive peak around zero is essentially related to the discretization of the power data, i.e. the limited resolution of the time series $P(t)$. From this observations, it can be concluded that the investigated power performance dynamics does not follow a Langevin process in a strict sense. Consequently, a respective reconstruction is definitely only an approximation. This approximation may be sufficient to comprehend the drift dynamics (cf. Chapter 2) since the reconstruction of the drift dynamics solely requires a vanishing average of the noise process but not the δ -correlation and the Gaussian distribution – as directly deduced from the respective Langevin equation. The reconstruction of the diffusion coefficient, on the other side, is essentially affected by the missing properties. This also has to be considered for the third kind of self-consistency test we present in the following paragraph.

A third alternative to test the self-consistency of the obtained results for the power characteristic is to simulate a new data set on the basis of the reconstructed

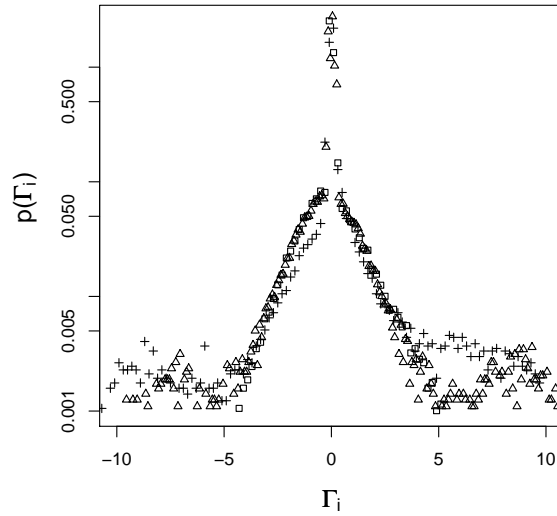


Figure 3.21: The pdfs of the reconstructed noise processes $\Gamma_i(t)$ for the wind speed bins $u_i = (6.25 \pm 0.25)\text{m/s}$ (\square), $u_i = (9.25 \pm 0.25)\text{m/s}$ (\triangle), and $u_i = (12.25 \pm 0.25)\text{m/s}$ (+). The resulting time series consist of approximately $104 \cdot 10^3$, $64 \cdot 10^3$ and $28 \cdot 10^3$ data points, respectively.

drift and diffusion coefficients, and compare the properties of this data set with those for the empirical data. In detail, this means to integrate (3.15) with the estimated values for $D_i^{(1)}(P_j)$ and $D_i^{(2)}(P_j)$ and a Gaussian distributed and δ -correlated random process $\Gamma_i(t)$ with zero mean and variance two. I performed this reconstruction for the considered set of measured wind speed and power output data. From the reconstructed data, I derived the increment time series $P_\tau(t)$ for different values of τ and compared the corresponding pdfs $p(P_\tau)$ with those for the empirical increments that were already analyzed in section 3.2.3. The results for four different values of τ from 1 s to 30 s are shown in figures 3.22(a) and (b). It is obvious that the model does not reproduce the bump structures of the empirical data. This has been in line with our expectations since these structures are due to turbine- or setup-specific effects that are not captured by the proposed model. Furthermore, the reconstructed increments are symmetrically distributed around zero whereas the empirical increments have an asymmetric pdf. Apart from these specific structures, the width of the pdfs for the reconstructed increments increases much faster with τ than that for the empirical data. Accordingly large is the deviation in the width of the two pdfs already for $\tau = 30$ s (see figure 3.22(b)). Instead of comparing the increment distributions, it is also possible to compare directly the pdfs of the measured and the reconstructed power data $P(t)$ and in particular their mean values. The result is quite similar – the pdfs agree roughly but not perfectly, and specific structures are not reconstructed by means of the stochastic model (cf. [Anahua et al 2008]).

I complete the discussion of the self-consistency of our model with the following conclusions. On the basis of numerical data, I have shown that the proposed scheme allows for an appropriate reconstruction of the theoretical power curve with respect to that the data has been simulated. Regarding the application to measured data, I maintain that the noise process separated from the analyzed empirical data does not correspond to Gaussian distributed noise, and a data reconstruction applying such an ideal noise process does not give results that perfectly agree with the empirical

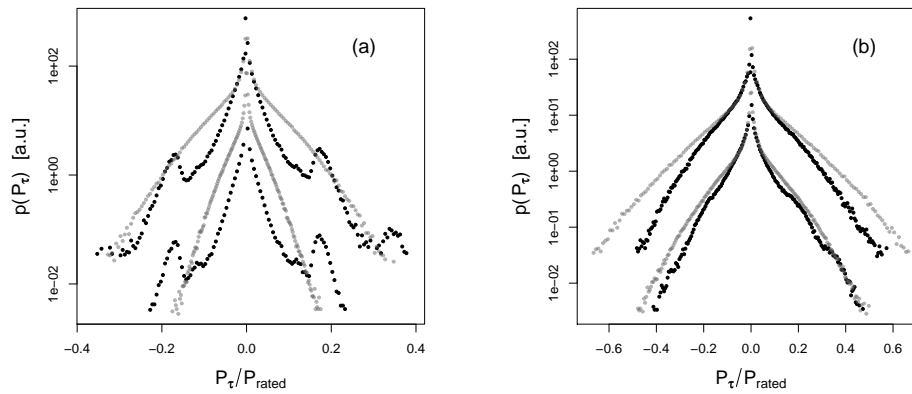


Figure 3.22: Pdfs of the empirical (black) and reconstructed (grey) increments $P_\tau(t)$ for different values of τ – (a) $\tau = 1$ s and $\tau = 5$ s, (b) $\tau = 15$ s and $\tau = 30$ s (from the bottom up in both cases). The upper pdfs are shifted in vertical direction for clarity of presentation. For the reconstruction we derived drift and diffusion coefficients according to (3.18–3.19) and inserted them into (3.15).

data. The latter was shown for the increments derived for the measured as well as for the reconstructed data of power output. Thereby, I have deduce that the pdfs for the reconstructed and the empirical increments agree to a varying degree for different values of the time increment τ . That is, the quality of the reconstruction depends on the considered time scale. At the same time, it is obvious that to model a time series not only on a specific scale but for a range of scales the simple procedure I performed here does not suffice and a multi-scale approach is necessary (cf. [Nawroth & Peinke 2006]).

3.4.2 Time scale of description

Following up the preceding point of discussion, the question arises to what extent the specification of a certain time scale is of importance for the modelling and reconstruction of the power performance dynamics. The dynamical model (3.15) is set up as a response model for the power output with regard to the wind speed fluctuations. It is however not clear on which time scale this response actually occurs. In addition, wind speed and power output, defining together the power performance, are given by two separately sampled time series that may correspond to two different time scales of description. Considering only the time series of the power output and its sampling rate as decisive time scale, one comes to the following conclusion. If a too large time scale of description is chosen, important information on the dynamics may be ignored. A too small scale, on the other side, is related to an increased noise level but is, in general¹, not a problem for the Langevin approach since the additional noise is separated with the stochastic diffusion component – a kind of oversampling is actually required to apply the proposed reconstruction scheme. The fluctuations of the wind speed affect the modelling of the power output dynamics by means of the wind speed binning, that is introduced for the dynamical power performance modelling in order to overcome the problem of the non-stationarity of the data. A higher variability of the wind speed induces a more

¹except for processes that are characterized by a finite Markov length scale that defines a lower limit in this regard (cf. Chapter 1)

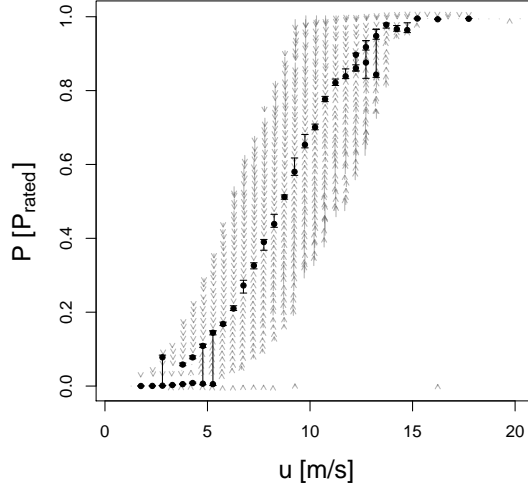


Figure 3.23: Dynamical power characteristic for the averaged wind speed \tilde{u}_T ($T = 60$ s) – the width of the power bins is $P_{rated}/60$. The arrows denote the drift field defined by the values $D_i^{(1)}(P_j)$ for each wind speed bin u_i and each power bin P_j , and the dots give the (stable) fixed points of the relaxation dynamics.

frequent switching between the single wind speed bins. Applying a kind of filter to the measured wind speed data, thus, affects the partitioning of the power output data with respect to the different wind speed bins.

That this may affect the results for the reconstructed power characteristic is demonstrated with the following approach. I applied a two-sided moving average to the wind speed data $u(t)$ to obtain the time series $\tilde{u}_T(t)$, i.e.

$$\tilde{u}_T(t) := \frac{1}{T+1} \sum_{\tau=-T/2}^{T/2} u(t-\tau), \quad (3.23)$$

using different values for the size of the averaging window T . The moving average acts as a low-pass filter on the wind speed fluctuations, and the size of the window defines a respective time scale.

I derived the dynamical power characteristic for the time series $\tilde{u}_T(t)$ and $P(t)$. The results for a window size of $T = 60$ s are shown in figure 3.23. The resulting dynamical power curve is considerably smoother than that for the non-averaged wind speed time series in figure 3.19(b) – for both results I used the same power binning. Likewise, the overall uncertainty of the power curve is smaller when considering the averaged wind speed. Adding up the individual errors for all obtained fixed points, I have obtained to a value of 2.84 for the power curve in figure 3.19(b) and only 1.09 for that in figure 3.23 (both in units of P_{rated}). With respect to the single fixed points, this gives an averaged value of 0.056 and 0.028 per reconstructed fixed point, respectively.

In [Wagner et al 2008] the quality of a certain power curve is determined as its total or alternatively its average uncertainty that is defined by the standard deviations for the points of the curve in the single wind speed bins, i.e. the values I determined above. Adopting this definition of evaluation, it can be concluded that the introduction of a moving average for the wind speed improves the results for the dynamical power characteristic – at least for the specific data set we analyzed.

The specification of a respective averaging window T is however to a large extent arbitrary. I propose to include it in the definition of the reconstruction procedure and fix it for a specific type of WECS and a particular measurement equipment and setup. It is important to note at this point that we do not apply the moving average to the power output data but solely to the time series of wind speed. The data basis with respect to that the drift coefficient is derived remains the same but is arranged in a modified order.

3.4.3 Representativity of considered wind speed

In the previous section, I discussed how the wind speed, that is considered for the description of a wind turbine's power performance, is best defined with respect to an appropriate time scale. Following these descriptions, I consider in this section two different spatial aspects that may influence the representativity of the horizontal wind speed $u(t)$ for the power performance modelling. A limitation of the representativity may hereby either imply that the time series $u(t)$ is systematically incorrect with respect to the simultaneously measured power output, i.e. it reflects a wind velocity that is systematically lower or higher than the wind velocity actually acting on the wind turbine, or that there is a lack of correlation between $u(t)$ and the corresponding time series for the power output $P(t)$. I address both issues here. As first effect, I discuss the delay between the wind and the power measurement that corresponds to a displacement in horizontal direction. The second aspect is the effect of wind shear that causes an inaccuracy with respect to the vertical direction of the wind field acting on the turbine. The section concludes with a general remark on secondary variables.

Delay effects

Probably, the most obvious disturbance, affecting the representativity of the wind speed time series $u(t)$ for the evaluation of power performance, is given by the spatial distance between the wind turbine and the wind measurement. Due to the IEC 61400-12-1 standard, the meteorological mast equipped with the anemometer may be between two and four times the rotor diameter away from the turbine, which results in travelling times of 30 s and longer dependent on the size of the turbine and the respective value of the wind speed. For this quantification, it is assumed that the turbulent structures do not change or dissipate over the travelling distance and that no new structures are formed – which is definitely a quite strong simplification. The non-vanishing travelling times directly cause a lack of correlation between $u(t)$ and $P(t)$. Applying an averaging procedure to the measured data with an averaging period that is to a considerable degree larger than the delays between wind and power measurement, the final results for the power curve are not affected by this lack of correlation. This argument may apply to the standard procedure due to IEC 61400-12-1 and similar approaches that are based on averages over periods of at least 10 min. For a reconstruction procedure that is directly based on the high-frequency data, however, this is not valid.

To evaluate the influence of the delay effects on the stochastic power performance model, I studied the relation between the relaxation and the correlation between $u(t)$ and $P(t)$ in more detail. The relaxation, defined by the drift term in the Langevin equation for each wind speed bin, causes a shifting of the maximum correlation between wind speed and power output – thus, the power output is delayed in comparison to the wind speed, and the respective time lag is the larger the lower the relaxation, i.e. the smaller e.g. the absolute value of the constant α_i in the linear approximation (3.16) is. On the other hand, an additional delay caused by the distance between wind measurement and turbine is reflected in a modified

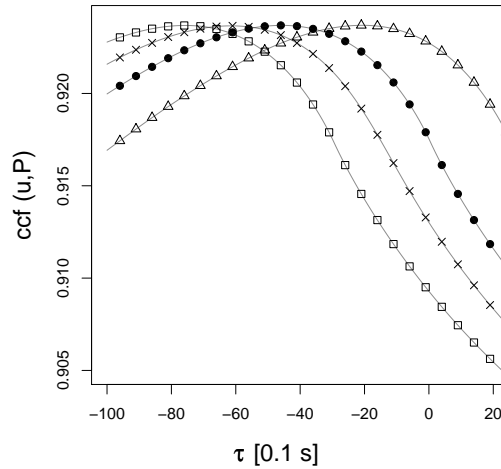


Figure 3.24: Correlation (defined as normalized cross correlation function) between wind speed and power output time series for numerical data simulated on basis of linear relaxation model ($\alpha_i=0.1$, $\beta_i=0.01$) according to (3.15–3.17) without additional delay (\bullet), a delay of $\delta = 30$ time steps (\square), $\delta = -25$ time steps (\triangle) and a random delay with a mean of $\bar{\delta} = 15$ and a standard deviation of $\sigma_\delta = 7$ time steps (\times).

drift term that is reconstructed with the procedure of stochastic modelling. This is illustrated in figures 3.24–3.26, showing the correlation (defined as a normalized cross correlation function) between wind speed and power output time series, the reconstructed drift coefficient for an exemplary wind speed bin and the resulting fixed points for all bins, derived for numerical data that is simulated on the basis of the stochastic relaxation model (3.15–3.17), as it was already used above, with and without additional delay between wind speed and power data. I performed this analysis for a delay of $\delta = 30$ time steps, i.e. the wind speed is measured before the power output, $\delta = -25$ time steps, i.e. the wind speed is measured after the the power output, and for a random delay. For the latter, I generated a uniformly distributed delay variable and smoothed it by a moving average with an averaging window of 30 time steps, resulting in a Gaussian distributed variable with a mean of $\bar{\delta} = 15$ and a standard deviation of $\sigma_\delta = 7$ time steps. Figure 3.24 indicates that the effective delay between the two time series, determined by the point of maximum correlation, is the sum of the delay caused by the finite relaxation of the dynamical process and the subsequently added delay. Figures 3.25(a) and (b) illustrate how adding an extra delay, in addition to that caused by the relaxation, influences the reconstructed drift coefficient. Positive as well negative additional delays effect a decrease of the absolute values for the drift coefficient, the uncertainties of the reconstructed values hereby decrease. The reconstructed drift coefficient hence corresponds to an effective relaxation that is the result of a combination of the actual relaxation behaviour of the WECS and additional delays. Note that also the reconstructed values of the drift coefficient for the model without additional delay deviate from the theoretical function (see figure 3.25(a)). These deviations are supposed to be caused by the effects discussed in Chapter 2, primarily the finite (comparatively low) sampling rate the data is sampled with. The influences on the complete dynamical power curve, defined by the fixed points for the single wind speed bins, are shown figures 3.26(a) and (b). The estimation of the fixed points is largely not affected by the changes in the slopes of the drift coefficients – assuming

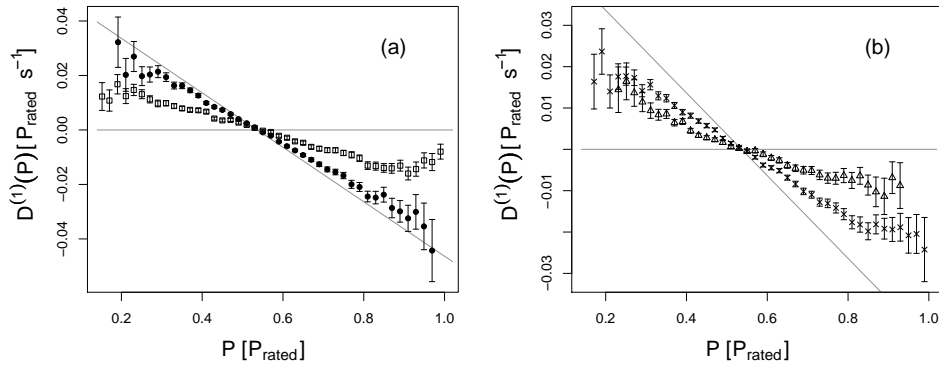


Figure 3.25: Reconstructed drift coefficients (for exemplary wind speed bin $u_i=(9.75\pm 0.25)\text{m/s}$) for simulated data of stochastic relaxation process with and without additional delays. For details see figure 3.24. The theoretical drift function, defined by the input value of α_i for the numerical model, is shown as decreasing solid line.

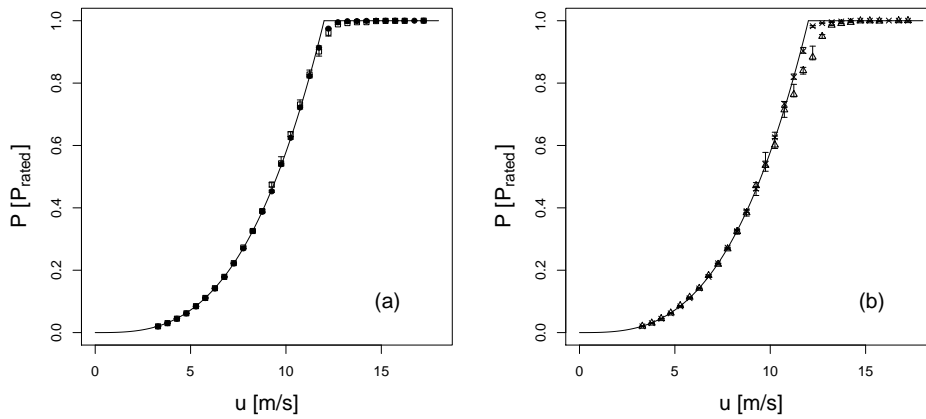


Figure 3.26: Reconstructed dynamical power curves for simulated data of stochastic relaxation process with and without additional delays. For details see figure 3.24.

that the delay caused by the spatial distance does not dominate the relaxation behaviour, and its magnitude is comparable to or smaller than the intrinsic relaxation. Solely for large negative delays, the results show significant deviations for the fixed points at the point of transition at rated wind speed (see curve denoted by triangles in figure 3.26(b)).

To conclude, the effect that additional delays in the measurement are included in an effective relaxation and approximately averaged out with respect to the fixed-point analysis, even though increasing the uncertainty of the results, serves as an essential criterion for the robustness of the reconstructed power characteristic. An explicit correction, e.g. by quantifying the delay and considering it in a dynamical delay equation (cf. section 1.4.), is not recommended against it. The delay does not correspond to a constant value but is a fluctuating variable determined by the short-term variations of wind speed and wind direction. Furthermore, these structures largely alter over the distance between anemometer and wind turbine. Therefore, it is obvious that it is unfeasible to quantify the actual short-term dynamics of the delay variable and directly incorporate it in the dynamical relaxation model.

Wind shear and other secondary variables

Another effect that may have a significant impact on the power performance measurement is the wind shear in vertical direction of the flow field, i.e. a systematic difference in the wind speeds for different heights. Commonly, a logarithmic wind profile is supposed to describe the dependence of the wind speed from the height above the surface. A consequence from this is that the main contribution of the electrical power produced by the WECS stems from the atmospheric layer that lies above the hub height of the turbine, which in turn also influences the correlation between measured wind speed and power output [Ragwitz 2001]. This argument raises the question how representative the wind speed measured at hub height as only variable actually is. In [Wagner et al 2008] it is proposed to consider instead an effective wind speed that is derived as an average of several wind speed time series measured at different heights and weighted by the area of the corresponding segment of the rotor swept area. In particular, it is shown that an additional measurement of the wind speed above the hub height of the turbine definitely decreases the overall uncertainty of a power curve derived on the basis of 10 min averages in accordance with the standard IEC 61400-12-1. Due to a lack of data at a sufficient number of different heights, I could not reconsider this investigations for our dynamical approach.

It has to be noted that a simultaneous measurement of wind speeds with anemometers at different heights is not common practice for power performance testing yet, and probably too expensive to be a standard. A promising approach is the utilization of LIDAR systems that are currently under investigation (see [Wagner et al 2008] but also e.g. [Wächter et al 2008]).

In general, there are two ways to include secondary variables as for example the wind shear in a reconstruction procedure for the power curve. On the one hand, they can be handled as additional parameters and by introducing a further binning. This however requires a data basis that increases with the number of bins for the additional parameter as factor of multiplication. The alternative is to include them in the definition of an effective wind speed. This is done with the air density in the standard 61400-12-1 in terms of a data normalization and the averaging over different heights in [Wagner et al 2008]. In order to introduce relevant secondary variables to the dynamical power characteristic proposed in this thesis, I definitely argue also for this second alternative.

There are other influences that significantly affect the power performance of a wind turbine but for that none of both approaches is recommended – examples are certain contaminations of the rotor blades, the effect of rain or icing. Situations where these disturbances appear should rather be excluded from the data basis for the analysis.

3.4.4 Relevance of the dynamical power characteristic

This section is concluded with a comment on the relevance of the dynamical power characteristic, summarizing, at the same time, the points of discussion listed in the preceding subsections. Based on the analysis of numerical data, it has been argued that the proposed dynamical approach gives more accurate results for the power curve than the standard procedure according to IEC 61400-12-1 (cf. [Anahua et al 2008, Gottschall & Peinke 2007, Gottschall & Peinke 2008a]). The results are however of a different kind. The resulting power characteristic corresponds to a dynamical model that contains considerably more information than solely the steady states of power performance. The reconstructed fixed points are defined through their attractiveness but this needn't imply that they actually define the states of average performance. For an asymmetric drift function, for instance, it

is obvious that mean value and fixed point are not identical. To estimate a quantity as the annual energy production (AEP), consequently, the values of the dynamical power curve should not be simply multiplied with the wind speed distribution as in (3.3), but one rather should utilize the reconstructed effective dynamical equations defined by drift and diffusion coefficients to simulate power output data on the basis of a reference wind speed time series.

On the other hand, it has also been shown that the set of reconstructed Langevin equations is only an approximation of the actual process dynamics – in particular, the included noise process does not correspond to a well-defined Langevin force which implies that the diffusion part of the dynamics is not reconstructed in an appropriate way. Consequently, a simulation of power output data on the basis of the estimated effective equations, including drift and diffusion, is not expected to give reliable results. Since the final estimation of the fixed points is largely not affected by the discussed influences, the resulting power characteristic is nonetheless a promising tool to identify systematic relative changes in the power performance, and therefore may be utilized for a monitoring analysis. Hereby, it should be kept in mind that the relevance of the complete power characteristic is once more restricted by the fact that the reconstruction procedure requires several definitions, as e.g. the fitting range for the estimation of the drift and diffusion coefficients or the definition of the considered wind speed, that are determined in an more or less arbitrary way. Only an appropriately specified determination of these criteria ensures the respective reproducibility.

In section 3.1.3, the introduction of a dynamical approach for the estimation of a power characteristic has been motivated with reference to several dynamic effects that must be considered when describing the power performance of a WECS on small time scales. Some of them are covered by the simple dynamical relaxation model, we proposed in 3.3.1 and discussed in this chapter, but not all of them. The deterministic relaxation in the introduced set of Langevin equations models the dynamic response behaviour of the WECS including the control dynamics. As an effective relaxation it also incorporates the delay effects caused by a spatial distance between turbine and wind measurement that do however not influence the estimation of the fixed points for the most part, as stated above. Not covered are, on the other side, systematic deviations for the reconstructed power performance that may e.g. result from wind shear effects that induce an asymmetric distribution of the wind speeds over the rotor area in vertical direction. Similar as for the standard procedure according to IEC 61400-12-1, it must be discussed if these effects should be considered by means of the definition of an effective wind speed.

3.5 General potential of a stochastic fixed-point analysis

The reconstruction of the dynamical power curve is based on the assumption that the fixed points of the deterministic dynamics serve as characteristic points for the described process. One of the most significant advantages of the definition of a fixed point as steady state in comparison e.g. to a mean value or maximum, that similarly may characterize a system and are more straightforward to determine, is its flexibility. With flexibility I refer to two different aspects – the applicability of the procedure to processes with multistable and asymmetric dynamics, respectively.

In particular, the definition of the fixed point is not restricted to processes with only one stable state but can be applied to systems with more than one stable fixed point in almost the same manner. This is demonstrated in figures 3.27(a–b) that show the power characteristic and the dynamical power curve as well as the corresponding curve due to the standard IEC 61400-12-1 for a MW-class turbine with a multistable behaviour in the region of transition at the value of rated wind speed (cf. [Gottschall & Peinke 2008b]). It is evident that the reconstruction procedure due to the IEC standard cannot detect more than one characteristic state per wind speed bin, and thus averages the multistability out (see figure 3.27(b)).

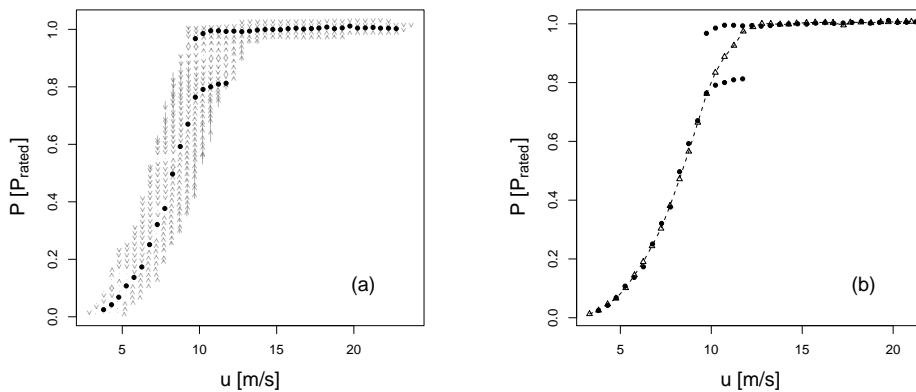


Figure 3.27: (a) Dynamical power characteristic for an exemplary MW-class turbine with multistable dynamics in the range of transition at rated wind speed. The arrows denote the drift field defined by the values $D_i^{(1)}(P_j)$, and the dots give the stable fixed points of the relaxation dynamics. (b) Dynamical power curve, composed of stable fixed points, in comparison to results due to IEC 61400-12-1 standard (triangles and dashed line).

Also for a process that is characterized by only one fixed point, the flexibility of a stochastic fixed-point analysis is of particular importance. This particularly becomes obvious for a process with asymmetric fluctuations. For the example of the power performance of a wind turbine, asymmetric power fluctuations are caused by a nonlinear transfer of wind fluctuations to the power output – i.e. the nonlinearities involve the asymmetries. The consequence is that a mean value is not well defined for the resulting asymmetric process and affected by systematic deviations (cf. [Gottschall & Peinke 2008a] and Appendix A, resp.).

A quite similar application, including such a nonlinear relaxation, is related to the performance of a cup anemometer (cf. [Hölzer 2008]). The response dynamics of a cup anemometer to incoming wind fluctuations is basically described by an asymmetric dynamical process. The response behaviour to an increase in wind speed is characterized by a faster relaxation than that to a decrease in wind speed – this is also referred to as overspeeding effect. A simple averaging of the high-

frequency signals from the cup anemometer thus results in an overestimation of the actual wind speed. The need for considering the specific properties of this dynamic response for free-wind conditions, in comparison to the static response measured in typical calibration procedures performed in a wind tunnel, has already been stated in [Antoniou et al 2001]. The dynamical approach, based on the reconstruction of an effective Langevin equation, may provide the basis for an advanced calibration procedure. Transferring the concept of a characteristic curve from the dynamical power characteristic to this application, it is straightforward to define an appropriate calibration procedure. The measured wind speed can be binned according to a reference wind speed and related to the fixed point for the respective data set in each bin. This is again a quasi-onedimensional approach that gives a two-dimensional calibration characteristic.

A crucial point for the application of the stochastic fixed-point analysis is the robustness of the derived fixed points. In Chapter 2, the influence of a finite resolution of the analyzed data as well as the presence of external measurement noise on the reconstruction of the fixed points of the respective underlying process has been discussed. Both investigated disturbances account for the problem that the reconstructed values for the drift coefficient are initially only estimations before applying a more complex optimization procedure. The quality of the estimates for the drift coefficient directly influences the robustness of the respective fixed point. In this chapter, these disturbances have been neglected so far since the fixed point is much less affected by them than the values for the drift functions. Respective deviations for the fixed points are supposed to be not significant in comparison to their total uncertainty. A more detailed investigation is recommended for the future.

As last issue, I address the convergence of the fixed points in relation to the number of considered data points and compare this behaviour with that for a respective mean value. For this, I again refer to the measured power performance data that was used for the power performance analysis and is described in Appendix B. Figure 3.28 shows two kinds of averages and two different definitions of fixed points for the power output as function of considered 10 min intervals of data for a specified wind speed bin. Mean values were calculated as bin averages of the values averaged over periods of 10 min according to IEC 61400-12-1, and as direct bin averages of the high-frequency data (circles and crosses, resp., in figure 3.28) – accordingly, the data was binned with respect to the 10 min averages of wind speed for the first case, and for the second with respect to the highly sampled wind speed data. The estimated fixed points differ in the point if the actually measured wind speed (full dots) is considered as binning parameter or a respective moving average (triangles). Apart from the systematic deviations, that are caused by asymmetries in the fluctuations, as discussed above, and handled differently by the single approaches, the studied quantities show also a different convergence behaviour. The bin average of the 10 min mean values as well as the fixed point for the averaged wind speed time series as reference, converge faster to a certain value and show afterwards less fluctuations than the two other quantities. For the specific convergence behaviour, it is thus of essential importance with respect to which wind speed data the corresponding data for the power output is binned. To achieve a good, i.e. a fast and robust convergence, it is recommended to filter the wind speed data with respect to short-term fluctuations, as done with the moving average. If a fixed point generally converges faster or possibly less fast than a respective mean value is hard to say after this analysis. The most relevant aspect, at least for this issue and this application, is the way the non-stationarity of the wind speed is dealt with.

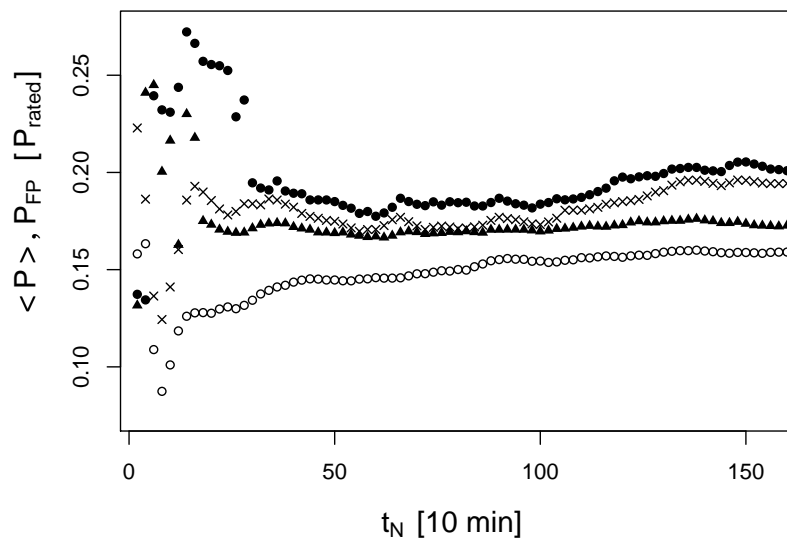


Figure 3.28: Convergence behaviour of mean values and fixed points for the power output in relation to the number of considered data points for the wind speed bin $u_i = (5.75 \pm 0.25)$ m/s. Mean values are derived as bin averages of 10 min mean values (open circles) and bin averages of the high-frequency data (crosses), fixed points with respect to the actually measured wind speed time series (full circles) and a respective moving average (full triangles). The number of considered data points is given as number of 10 min intervals or as sets of 600 data points, respectively. For a description of the used data see Appendix B.

3.6 Appendix A: How to improve the estimation of power curves for wind turbines ²

We introduce a dynamical approach for the determination of power curves for wind turbines and compare it with two common methods – among them the standard procedure due to IEC 61400-12-1, i.e. the international standard prepared and published by the International Electrotechnical Commission. The main idea of the new method is to separate the dynamics of a wind turbine into a deterministic and a stochastic part, corresponding to the actual behaviour of the wind turbine and external influences such as the turbulence of the wind, respectively. Particularly, the governing coefficients are reconstructed from the data, and the power characteristic is extracted as the steady states of the deterministic behaviour. Our results suggest that a dynamical approach enables to grasp the actual dynamics of a wind turbine and to gain most accurate results for the power curve, independent of site-specific influences.

A.1. Introduction

With increasing importance of wind as sustainable energy source and, consequently, the spreading of wind turbines as wind energy converter systems (WECS), wind energy research has become a new orientation. Physical findings have to be combined with technical know-how. An improved understanding of the performance of a WECS, as well as of e.g. the loads affecting certain parts of a wind turbine, can only be achieved if we comprehend its fuel, the wind, in more details.

In the centre of our research stands the assessment of the wind turbine's performance, i.e. the description of the wind energy conversion process. The so-called power curve, the electrical power output versus the wind speed, summarizes the technical characteristics of the whole turbine and is approved to be the most important testimony for the wind turbine's performance in the view of the operator. The shape of a power curve is governed by the cubic relation between the wind speed and the corresponding power in the wind up to the so-called rated wind speed where the power takes a constant value until the turbine cuts out. The actual energy density of the wind is reduced by the energy not usable due to a physical limit and additional losses due to several technical characteristics. (See e.g. [Hau 2005].) The correct and effective determination of these effects is of central interest.

A standard method for the determination of wind turbine power curves is given by [IEC 2005a], referred to as IEC 61400-12-1. This norm is an averaging procedure, easily applicable if enough data is available. The resulting IEC curve is the common tool to estimate a wind turbine's energy yield. However, it cannot be used to display e.g. the short-time fluctuations of the electrical power output induced by turbulent wind conditions or to explain the orographic dependencies of a wind turbine's performance.

Recently, an alternative method for the estimation of wind power characteristics has been established in [Anahua et al 2007, Anahua et al 2008]. The main idea is to reconstruct the short-time dynamics of the power conversion process and to determine the power curve as the steady states or fixed points of this process – these results represent nothing else than the ideal performance for non-fluctuating laminar wind conditions. Therefore, we have introduced the term dynamical power curve. In detail, the dynamics of the WECS is considered to be a Markovian stochastic process, and the wind speed is interpreted as a noisy driving force. The

²Published in a slightly modified version as J. GOTTSCHALL and J. PEINKE: How to improve the estimation of power curves for wind turbines. In: *Environ. Res Lett.* 3, 015005 (2008).

crucial point of this method is to divide the dynamics into a deterministic and a stochastic part. The stochastic part, given by dynamical noise, summarizes all the otherwise unseizable microscopic interactions and enables a macroscopic description of the considered open complex system. I.e. it takes also into account that the scalar wind speed, measured at a meteorological mast, cannot represent the complete wind field actually acting on the wind turbine. Following Haken [Haken 1983], this can be regarded as a synergetic approach to complexity. For details about the general method and other applications see e.g. [Friedrich et al 2000].

The aim of this paper is to show the advantages of the recently introduced method of the dynamical power curve estimation, comparing it with two other determination methods, namely the common IEC standard and the maximum principle proposed in [Rauh et al 2007], and going back to certain examples of measured and simulated data. Therefore, we first introduce the three considered methods in more detail, performing then the comparison of methods and conclude with a short discussion.

A.2. Dynamical power curve

The dynamical power curve approach is based on highly sampled data in order to reconstruct the actual process dynamics on small timescales – for more details on the short-term dynamics of the conversion process see [Gottschall & Peinke 2007]. Its basic idea is to describe the electrical power output P of the wind turbine as a diffusion process, i.e. a stochastic process that satisfies the Markovian property and that can be separated into a drift and a diffusion part. Then, a typical time series can be presented as $P(t) = P_{\text{FP}}(u) + p(t)$, where P_{FP} denotes a steady power value dependent on the wind speed u (or rather a non-fluctuating steady state or mean value of u , not further specified here), and $p(t)$ refers to the corresponding short-term fluctuations around this value caused by the wind turbulences and the response of the WECS to these. The time series $P(t)$ is assumed to be stationary with respect to a certain wind speed interval. Therefore, the dynamics of $P(t)$ is analyzed for each selected wind speed interval or bin separately. (Note that the mapping of the power values $P(t)$ to the single wind speed bins is defined by the actual wind speed $u(t)$. A split-up of the wind speed into a mean value and short-term fluctuations or another kind of filtering is not considered here. Fluctuations are analyzed in terms of increments. In the end, the reconstructed value of P_{FP} is mapped to the average of all values $u(t)$ lying in the corresponding bin.)

For the evolution in time of the variable P we formulate a set of Langevin equations

$$\frac{d}{dt}P(t) = D_i^{(1)}(P) + \sqrt{D_i^{(2)}(P)} \Gamma_i(t) \quad (3.24)$$

where the index i refers to the wind speed bin defined by u_i . $D_i^{(1)}$ is called drift coefficient and represents the deterministic part of the process, whereas the diffusion coefficient $D_i^{(2)}$ together with the Langevin force $\Gamma_i(t)$ representing δ -correlated Gaussian white noise ($\langle \Gamma_i(t) \rangle = 0$ and $\langle \Gamma_i(t_1) \Gamma_j(t_2) \rangle = 2\delta_{ij}\delta(t_1 - t_2)$) describes its stochastic part. The units of $D_i^{(1)}$ and $D_i^{(2)}$ are $[P]s^{-1}$ and $[P]s^{-2}$, respectively, with $[P]$ as the unit of P . The unit of $\Gamma_i(t)$ arises from the relation $\Gamma(t)dt = dW(t)$ and $dW(t)^2 = dt$ where $dW(t)$ is a Wiener process (see [Gardiner 1985]). A simple relaxation model, as proposed in [Rauh et al 2007], would follow the equation

$$\frac{d}{dt}P(t) = -\alpha_i [P(t) - P_{\text{FP}}(u_i)] + \sqrt{\beta_i} \Gamma_i(t), \quad (3.25)$$

where α_i is a constant relaxation factor that quantizes a relaxation around the steady state P_{FP} driven by the wind speed variation, and β_i the strength of additive

dynamical noise. The dynamical equation is set up for each wind speed bin u_i separately.

A reconstruction of this dynamics enables the estimation of the values $P_{\text{FP}}(u_i)$ and with it the determination of the power characteristic. Following [Risken 1989], the coefficients $D_i^{(n)}$ for $n = 1, 2$ are given by the conditional moments $M_{ij}^{(n)}$ that can be directly calculated from the data according to

$$D_i^{(n)}(P_j) = \frac{1}{n!} \lim_{\tau \rightarrow 0} \frac{1}{\tau} M_{ij}^{(n)}(\tau) \quad (3.26)$$

$$\text{with } M_{ij}^{(n)}(\tau) := \langle [P(t+\tau) - P(t)]^n \rangle |_{P(t)=P_j} \quad (3.27)$$

where $\langle \dots \rangle$ denotes an ensemble average and $|_{P(t)=P_j}$ the condition that the stochastic variable $P(t)$ is in the bin P_j at time t . The binning for P , defined by the discrete values P_j , is introduced analogously to the wind speed binning.

For small and finite τ , we make the approximation

$$M_{ij}^{(n)}(\tau) \approx n! \tau D_i^{(n)}(P_j) + \mathcal{O}(\tau^2), \quad (3.28)$$

applying an Itô-Taylor series expansion. Depending on the process it might be necessary to consider further higher-order terms. An exact derivation is given in [Friedrich et al 2002].

If additional measurement noise is present, $D_i^{(n)}$ is best estimated by the extrapolation

$$D_i^{(n)}(P_j) = \frac{1}{n!} \frac{M_{ij}^{(n)}(\tau_2) - M_{ij}^{(n)}(\tau_1)}{\tau_2 - \tau_1} \quad (3.29)$$

for suitable τ_1 and τ_2 (see [Böttcher et al 2006]).

Finally, a deterministic fixed-point analysis for each wind speed bin u_i yields the steady states $P_{\text{FP}}(u_i)$, defined by $D_j^{(1)}(P) \equiv 0$, and with it the dynamical power curve.

A.3. Comparison with other determination methods

A.3.1. The IEC standard

In short, the standard method according to IEC 61400-12-1 [IEC 2005a] is defined by relating the averages of wind speed and power output over 10 min, i.e.

$$\langle u(t) \rangle_{10 \text{ min}} \mapsto \langle P(t) \rangle_{10 \text{ min}}, \quad (3.30)$$

and averaging in a second step all values lying in a wind speed bin of the width of normally 0.5 m/s. Already in [Albers & Hinsch 1996] it has been argued that this procedure does not account for the nonlinearity of the power characteristic. Expressing the wind speed as $u = \bar{u} + u'$ where \bar{u} is the mean value, i.e. $\bar{u} = \langle u(t) \rangle$, and u' the corresponding short-term fluctuations around this value with $\langle u'(t) \rangle = 0$, the power output P as a function of u can be expanded in the Taylor series

$$P(u) = P(\bar{u}) + \frac{\partial P(\bar{u})}{\partial u} u' + \frac{1}{2} \frac{\partial^2 P(\bar{u})}{\partial u^2} u'^2 + \mathcal{O}(u'^3). \quad (3.31)$$

Assuming that the fluctuations $u'(t)$ are symmetric around \bar{u} , and neglecting the terms $\mathcal{O}(u'^3)$, one obtains for the averaged power output

$$\langle P(u) \rangle = P(\bar{u}) + \frac{1}{2} \frac{\partial^2 P(\bar{u})}{\partial u^2} \sigma^2 \quad (3.32)$$

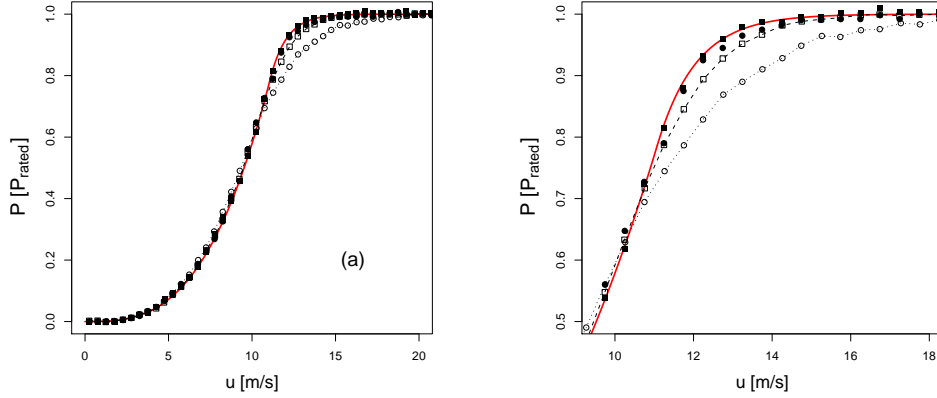


Figure 3.29: Reconstructed power curves for simulated data, (a) complete curve and (b) cutout, for different values of turbulence intensity. Results according to IEC standard are given by open symbols, stationary states according to dynamical method by full symbols (rectangle for $I = 0.10$ and circle for $I = 0.20$). The solid line (red) denotes the exact power curve, i.e. the input characteristic for our model.

with $\sigma^2 = \langle u'^2(t) \rangle$.

It follows $\langle P(u) \rangle \neq P(\langle u \rangle)$ for a nonlinear function $P(\bar{u})$ and non-vanishing σ . Defining the turbulence intensity $I = \sigma/\bar{u}$, one rather finds a correction term proportional to I^2 for $P(\bar{u})$ known. The inequality above indicates that symmetric wind speed fluctuations, as they are assumed, are transferred to asymmetric power fluctuations due to the nonlinearity of the power curve. It follows that a linear averaging procedure does not give exact results.

To illustrate this discrepancy, we simulated wind speed and power output data, using a simple relaxation model as given by (3.25) with typical parameter values we obtained from the analysis of measured data. We performed the simulation for different values of turbulence intensity, and reconstructed the power characteristic according to both the IEC standard procedure and the dynamical method. Results are shown in figure 3.29.

The systematic deviations of the results following the IEC recommendations from the real power curve, i.e. the input characteristic for our model, are evident. Even though the result of the dynamical method is not affected by the mentioned averaging problem, there is similarly a small discrepancy. This is due to a non-stationarity of the data induced by the predetermined wind speed binning, and increases with the turbulence intensity. However, it is much smaller than the errors of the IEC method, and for our set of measured data (see below) it is negligible.

A.3.2. Maximum principle according to Rauh

Rauh et al. proposed an even simpler method to determine the power curve of a wind turbine [Rauh et al 2007]. The idea is to define an empirical power curve by the location where, in a given wind speed bin, the maximal density of points $P(t_n)$ is found. I.e. the power curve is given by the points $\{u_i, P_{k(i)}\}$, where i is the number of the speed bin and $k(i)$ denotes the power bin with

$$N_k := \sum_n \Theta(P(t_n) - P_k) \Theta(u(t_n) - u_i) \quad \text{and} \quad N_{k(i)} \geq N_k, \quad (3.33)$$

where $\Theta(x)$ is a Heaviside function defined by

$$\Theta(x) = \begin{cases} 1 & \text{if } -\Delta/2 \leq x < \Delta/2 \\ 0 & \text{else.} \end{cases} \quad \text{with the particular bin width } \Delta \quad (3.34)$$

In words, we determine for each power bin k the number of events N_k , counting the points $P(t_n)$ lying in the respective bin. The largest number of points is denoted by $N_{k(i)}$, where $k(i)$ is the sought bin with the extremal property and, correspondingly, $P_{k(i)}$ the point of the power curve. (The values u_i and P_k can be defined either by the mid of the bin or by the mean value of the points lying in it.) Rauh et al. argue that this extremal property is expected if the power curve is an attractor.

To show the weakness of this method, we assume again that $P(t)$ can be described by a diffusion process and follows an equation as given by (3.24). As alternative to the Langevin equation, a diffusion process can also be described by a Fokker-Planck equation (see [Risken 1989])

$$\frac{\partial p(P, t)}{\partial t} = \left\{ -\frac{\partial}{\partial P} D^{(1)}(P) + \frac{\partial^2}{\partial P^2} D^{(2)}(P) \right\} p(P, t) = -\frac{\partial}{\partial P} S(P, t). \quad (3.35)$$

$D^{(1)}(P)$ and $D^{(2)}(P)$ are the coefficients defined in (3.26), the index i is omitted here. The specific kinds of description by a Fokker-Planck equation and a Langevin equation are equivalent – while the Langevin equation is a stochastic differential equation for the state variable and describes the actual trajectory of this variable, the Fokker-Planck equation is a partial differential equation for the probability density $p(P, t)$ of the state variable evolving in time. For $S = \text{const}$, we obtain the stationary solution

$$p_{\text{stat}}(P) = \frac{\mathcal{N}}{D^{(2)}(P)} \exp \left[\int^P \frac{D^{(1)}(\tilde{P})}{D^{(2)}(\tilde{P})} d\tilde{P} \right] \quad (3.36)$$

with the normalization constant \mathcal{N} .

The proceeding is now to investigate under which conditions the point of maximal density P_{max} is equal to the steady state P_{FP} , i.e. the dynamical power curve equals the power curve according to Rauh's maximum principle. For this purpose, we differentiate $p_{\text{stat}}(P)$ with respect to P and set it to zero. Since the exponential term is always > 0 and assuming that the lower integration term vanishes, we end up with

$$\frac{\mathcal{N}}{(D^{(2)}(P_{\text{max}}))^2} \left[D^{(1)}(P_{\text{max}}) - \frac{\partial}{\partial P} D^{(2)}(P_{\text{max}}) \right] = 0. \quad (3.37)$$

Because $D^{(2)}(P) > 0$ essentially, P_{max} has to fulfil $D^{(1)}(P_{\text{max}}) = \partial/\partial P \{ D^{(2)}(P_{\text{max}}) \}$. The definition of the fixed point provides $D^{(1)}(P_{\text{FP}}) \equiv 0$. That means that the point of maximal density P_{max} equals only the steady state if its derivative with respect to P vanishes, i.e. for $D^{(2)}(P) = \text{const}$ or if $D^{(2)}(P)$ has an extremum for P_{FP} . We call this type of dynamical noise ideal noise. Consequently, if a diffusion function $D^{(2)}$ does not satisfy this condition, the corresponding noise is called non-ideal.

The consequences for the reconstructed points of the power curve are, first, to be demonstrated with a one-dimensional example of simulated data, see figure 3.30. We integrated the Langevin equation for a process with $D^{(1)}(x) = -0.1x + \text{const}$, and a step function for $D^{(2)}(x)$ as realization of non-ideal noise. This behaviour is motivated by our observations for experimental data. The original functions as well as the reconstructed points, according to (3–7), are shown in figures 3.30(a) and (b). Examining the probability density function, shown in figure 3.30(c), we

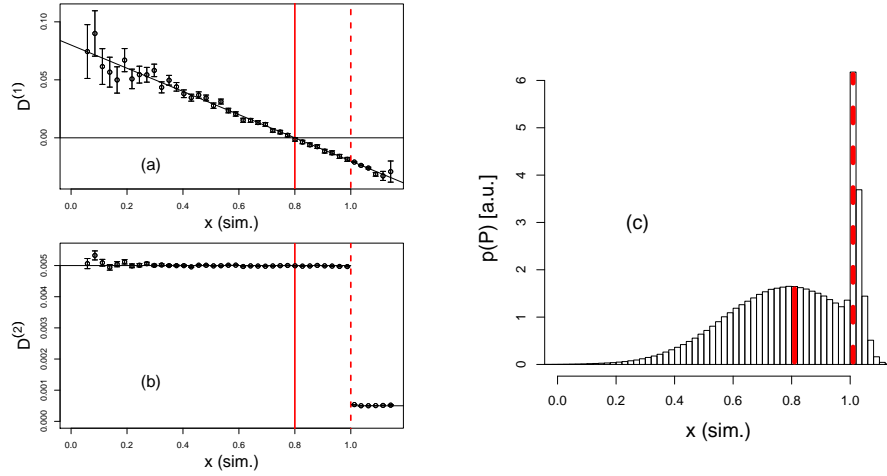


Figure 3.30: Simulated data and reconstruction of its dynamical coefficients to show the weakness of Rauh's maximum principle – (a) drift coefficient $D^{(1)}(x)$, (b) diffusion coefficient $D^{(2)}(x)$ and (c) histogram of data. The fixed point is marked with a solid line, the maximum with a dashed line, both in red.

find that the point of maximal density does not equal the actual fixed point of relaxation, but is affected essentially by the shape of $D^{(2)}$. The reconstruction of $D^{(1)}$, however, is not influenced and provides the correct result for the fixed point. Large errors of the reconstructed coefficients for small values of x are due a small amount of data (see histogram) – errors are proportional to $1/\sqrt{N}$, where N is the number of considered data points.

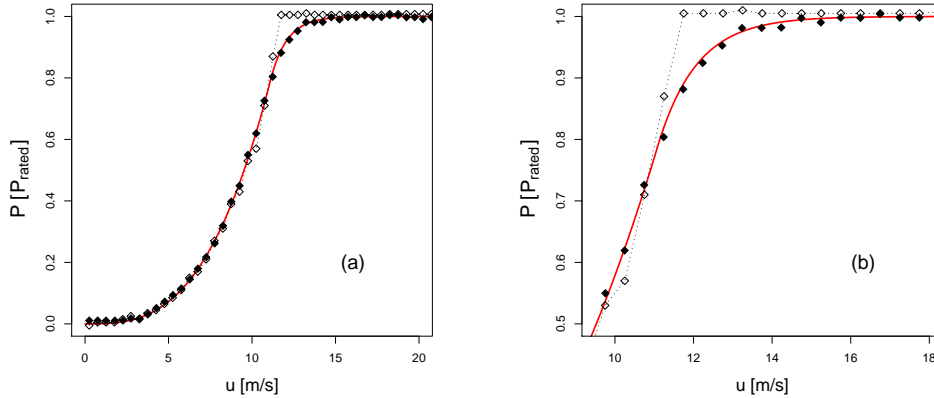


Figure 3.31: Reconstructed power curves for simulated data, (a) complete curve and (b) cutout. Results according to Rauh's maximum principle are given by open symbols and dotted line, stationary states according to dynamical method by full symbols. The solid line (red) denotes the exact power curve, i.e. the input characteristic for our model. The turbulence intensity of the simulated wind speed data is 0.10.

To transfer this one-dimensional example to a two-dimensional model for the power characteristic, we use again the relaxation model given by (3.25), and replace the constant diffusion coefficient β_i by a step function as shown in figure 3.30(b) with a decrease of diffusion strength to one tenth behind the step. As shown in

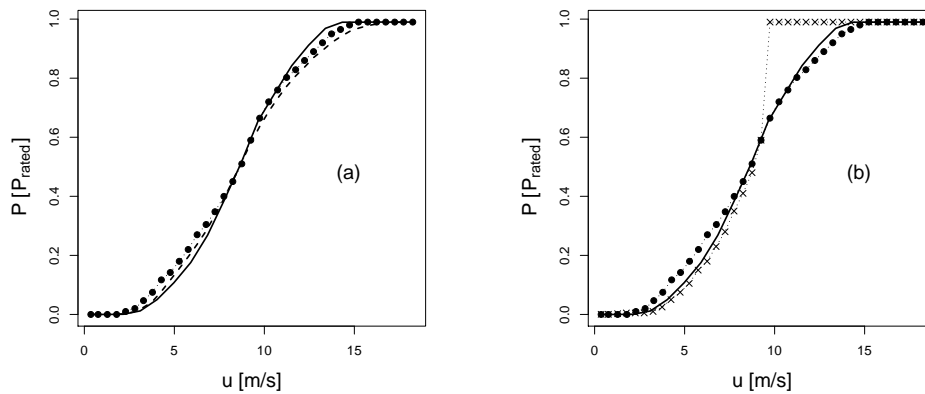


Figure 3.32: Schematic comparison of power curves due to IEC standard (solid line), reconstructed fixed points (full symbols) and (a) bin averages of 1 Hz data (dashed line), respectively, (b) maxima according Rauh's method (crosses) for measured data.

figures 3.31(a) and (b), Rauh's maximum principle overestimates the points of the power curve in the region of transition to the rated power, and seems overall to be less accurate than the dynamical method.

A.3.3. A look at real data

To show the actual impact of the effects explained in the two last subsections, we refer to a set of measured wind speed and power output data of a WECS. The data had been obtained from a commercial MW-class turbine, located in a wind park in the mid-western part of Germany. The wind speed data had been recorded by an ultrasonic anemometer that was placed on a meteorological mast in front of the turbine, satisfying the requirements stated in the IEC 61400-12-1 standard, with a frequency of 50 Hz. The analyzed data set is sampled with a frequency of 1 Hz, defined by the lower resolution for the recording of the power data. It consists of approximately 1,800,000 data points, which corresponds to a measurement period of three weeks. The averaged turbulence intensity of the wind speed data is 0.135.

At first, we compare the power curve estimated due to the IEC 61400-12-1 standard procedure with the dynamical power curve, i.e. the reconstructed fixed points. Applying both methods to the measured data, we do not observe the expected results according to 3.6 (see figure 3.32(a) in comparison to figure 3.29). The deviations are rather directly opposed – for small wind speeds and a positive curvature of the power curve the fixed points lie above the IEC curve, for large wind speeds up to the rated value and a negative curvature they lie below. In figure 3.32(a), we have additionally compared the IEC curve with a power curve that is obtained by simply averaging the high-frequency data in each wind speed bin instead of taking the average values over 10 min. The result of this alternative approach again shows deviations of a different kind. For low wind speeds as well as for high wind speed values, this high-frequency power curve lies below or respectively to the right of the dynamical power curve. In comparison to the IEC curve it is rotated in clockwise direction. To study these deviations in more detail, we analyzed the distributions of power output data for the single wind speed bins. In figures 3.33(a) and (b), we have compared the distribution of the original power data sampled by a frequency of 1 Hz to the distribution of the averages over 10 min (binned according to the corresponding wind speed averages) for two different wind speed bins. The his-

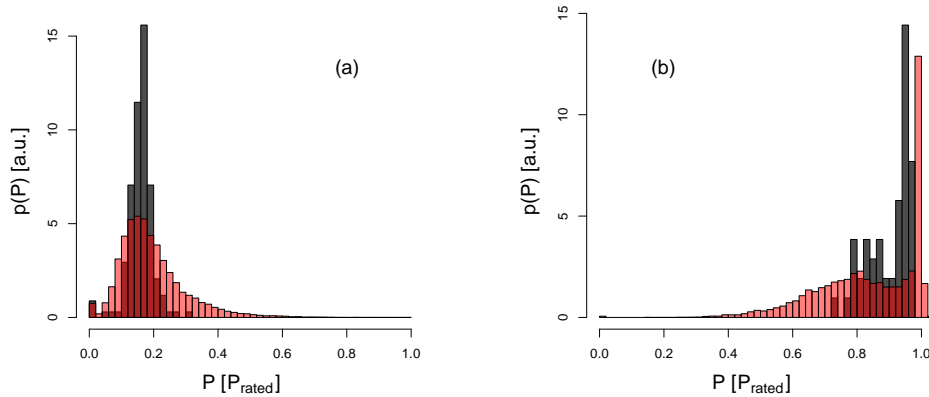


Figure 3.33: Histograms of power output data for the bins (a) $u = (5.75 \pm 0.25)\text{m/s}$ and (b) $u = (12.25 \pm 0.25)\text{m/s}$. Distribution of average values over 10 min. in dark grey, and of the high-frequency data (red) behind.

tograms for the high-frequency data are in both cases highly asymmetric. For low wind speeds, there is an additional peak at the left side of the spectrum, for high wind speeds at the right side – corresponding to the cut-in and rated values of the power curve. The averaging procedure seems to cut off the part of the other side of the distribution in each case. Hence, the mean value of the averages over 10 min is shifted to the peak value. This phenomenological approach clarifies the impact of asymmetries in the distributions of the power data that are primarily caused by the specific transfer dynamics that determine the power conversion. It does however not give an explanation of the specific behaviour observed in figure 3.32(a).

To inspect the application of Rauh’s maximum principle, we also refer to the probability distribution of the high-sampled power output data for each single wind speed bin, as exemplarily shown in figures 3.33(a) and (b). For wind speeds in the range of the rated value, the histograms have basically the same shape as the one for the simulated data in figure 3.30(c) (see figure 3.33(b)). The additional peaks result in a distinctive kink for the whole power curve, that is not present for the other approaches (see figure 3.32(b)). For the simulated data, the additional peak in the histogram has been traced back to non-ideal noise, implemented as a step function for the diffusion coefficient. Our observations let us suppose that similar dynamics underlie the measured data. An explanation for this specific diffusion dynamics are the control mechanisms of the wind turbine, especially at the point of transition to the rated power. We conclude that the system-specific behaviour of the WECS is not only present in the drift dynamics, as initially supposed, but also in the diffusion part of the process dynamics.

A.4. Conclusions and discussion

To describe the performance of a WECS in an appropriate manner, that is the main conclusion of our investigations, it is essential to grasp the actual dynamics of the process. As demonstrated, the location of both the maximum and the mean value is substantially influenced by the shape of the distribution of data. Simple sample statistics seems not to be sufficient to capture its specific characteristics, a detailed analysis of the underlying dynamics and a more flexible definition of a characteristic steady state is rather necessary. Therefore, we have introduced a so-called dynamical approach, estimating this steady state or fixed point by extracting the actual

deterministic dynamics of the wind turbine. Stochastic influences are handled as noise and separated from this information. For this reason, the dynamical approach is a generic one and promises more accurate results than the two other presented methods. Beyond, it provides much more information than only the steady states of the process, i.e. the power curve, and potentiates e.g. a detailed analysis of the short-term dynamics of the WECS (see [Gottschall & Peinke 2007]). Admittedly, the dynamical method is just for this reason of another type than the IEC standard and cannot replace it that easily. Instead of mean values we estimate fixed points, which has essential implications on the application of the resulting power curves.

3.7 Appendix B: Data description

The investigations presented in Chapter 3 are mainly based on a set of measured wind speed and power data that was recorded during a measurement campaign in a wind park in the mid-western part of Germany. The campaign lasted from 2004 until April 2006. Unless otherwise mentioned we refer to a data set that was recorded in the months October till December 2005, consisting of 58 complete but not necessarily successive days of data. The wind park is located near the village Meerhof, south of Paderborn, and the area is described as the foothills of a low mountain range. The site itself is largely flat, the distance of the forest area north of the measurement setup as well as of the neighbouring wind turbines is large enough not to consider them as factors for flow distortion (see figure 3.34(a)).

A schematic description of the measurement setup is given in figure 3.34(b). The meteorological mast was located in a distance of $D = 180$ m to the west of the considered turbine ($\phi_0 = 87^\circ$). With a rotor diameter of $d = 70$ m, this corresponds to a distance of roughly $2.5d$ between turbine and mast. The disturbed wind sector, corresponding to the wind directions excluded due to the wake of the wind turbine on the mast, has been identified as the interval $[40.5^\circ, 133.5^\circ]$. A wind rose indicating the main flow directions is shown in figure 3.35. The meteorological mast was equipped with several cup and ultrasonic anemometers, as well as further meteorological sensors at different heights. Thereof we only use the data of an ultrasonic anemometer of the type USA-1, manufactured by the company Metek, that was located at a height of 95 m which is 3 m under the hub height of the turbine.

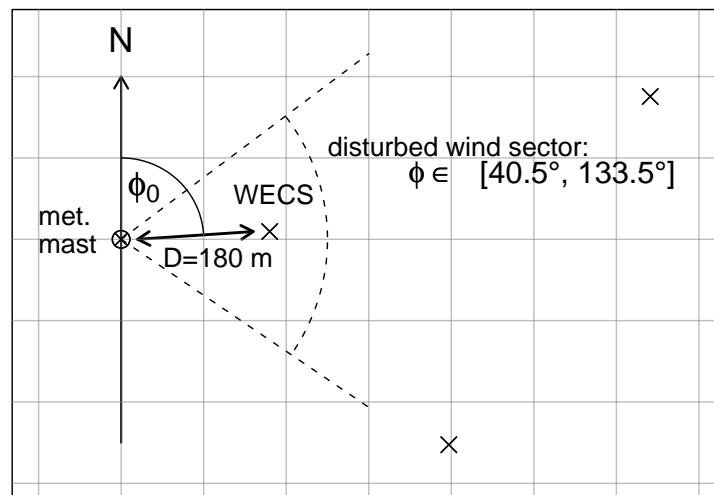
The ultrasonic anemometer consists of three ultrasonic transponder pairs that are mounted around a bar in the center. The middle bar causes a flow distortion that was accounted for by a correction algorithm applied subsequently to the measurement. The data, consisting of the three velocity components X , Y , Z for North-South, East-West and the vertical direction, was recorded with a sampling frequency of 50 Hz that was reduced to 1 Hz and 10 Hz, respectively, for the analysis by simple averaging. From the two horizontal components X and Y , the horizontal wind speed u , i.e. the absolute value of the horizontal wind velocity component in flow direction, and the wind direction ϕ were derived. A data normalization with regard to the air density, as required in [IEC 2005a], could not be applied due to a failure of the air pressure sensor.

The tested wind turbine is a commercial MW-class turbine with a hub height of 98 m. The power output data were sampled with a frequency of 1 Hz. For the analysis it was normalized to rated power.



(a)

Image Copyright 2008 GeoContent
Copyright 2008 Google



(b)

Figure 3.34: Site layout and location of the meteorological mast. Both figures are aligned in North-South direction. In (a), the considered turbine is the most-left one, i.e. at the west borderline of the wind park. (b) is a schematic cut-out of (a). D gives the distance between WECS and mast, and ϕ_0 the corresponding direction ($\phi_0 = 87^\circ$). Crosses indicate the considered and the neighbouring turbines.

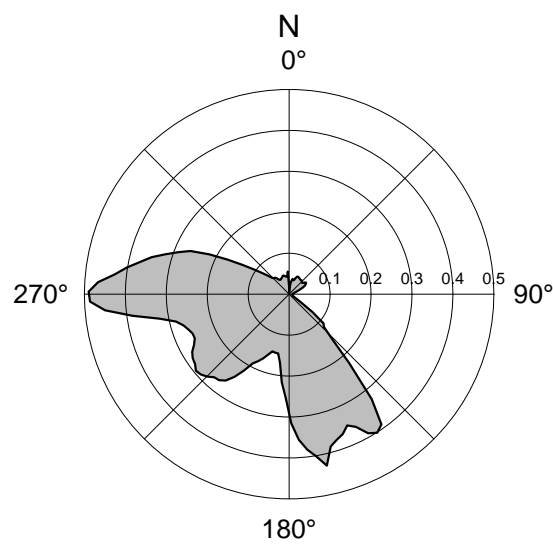


Figure 3.35: Distribution of wind directions for the considered site. Direction data extracted from the two horizontal components of an ultrasonic anemometer, with a sampling rate of 1 Hz. The scale of the polar plot gives the probability density of the wind directions. The analyzed data set consists of data from 58 complete but not necessarily successive days between October and December 2005.

Bibliography

- [Albers & Hinsch 1996] A. Albers and Ch. Hinsch: Influence of Different Meteorological Conditions on the Power Performance of Large WECs. In: *DEWI Mag.* 9, 40–49 (1996).
- [Albers et al 2007] A. Albers, T. Jakobi, R. Rohden, and J. Stoltenjohannes: Influence of Meteorological Variables on Measured Wind Turbine Power Curves. In: *Proc. of EWEC conference* (published online) (2007).
- [Anahua et al 2007] E. Anahua, St. Barth, and J. Peinke: Characterisation of the Power Curve for Wind Turbines by Stochastic Modelling. In: *Wind Energy – Proc. Euromech Colloquium* ed J. Peinke, P. Schaumann and St. Barth (Springer, Berlin), 173–177 (2007).
- [Anahua et al 2008] E. Anahua, St. Barth, and J. Peinke: Markovian Power Curves for Wind Turbines. In: *Wind Energy* 11, 219–232 (2007).
- [Antoniou et al 2001] I. Antoniou, F. Mouzakis, A. Albers, U. Follrichs, T. Curvers, H. Verhoef, P. Enevoldsen, J. Højstrup, and L. Christensen: Identification of variables for site calibration and power assessment in complex terrain (Project JOR3-CT98-0257). In: *Proc. of EWEC conference*, 17–22 (2001).
- [Beck 2004] Ch. Beck: Superstatistics in hydrodynamic turbulence. In: *Physica D* 193, 195–207 (2004).
- [Bianchi et al 2006] F. D. Bianchi, H. De Battista, and R. J. Mantz: *Wind Turbine Control Systems*. Springer, Berlin; 2nd edition (2006).
- [Böttcher et al 2003] F. Böttcher, Ch. Renner, H.-P. Waldl, and J. Peinke: On the statistics of wind gusts. In: *Boundary-Layer Meteorology* 108, 163–173 (2003).
- [Böttcher et al 2006] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase: Reconstruction of Complex Dynamical Systems Affected by Strong Measurement Noise. In: *Phys. Rev. Lett.* 97, 090603 (2006).
- [Böttcher et al 2007a] F. Böttcher, St. Barth, and J. Peinke: Small and large scale fluctuations in atmospheric wind speeds. In: *Stoch. Environ. Res. Ris. Assess.* 21, 299–308 (2007).
- [Böttcher et al 2007b] F. Böttcher, J. Peinke, D. Kleinhans, and R. Friedrich: Handling systems driven by different noise sources – Implications for power curve estimations. In: *Wind Energy – Proc. Euromech Colloquium* ed J. Peinke, P. Schaumann and St. Barth (Springer, Berlin), 179–182 (2007).
- [Burton et al 2001] T. Burton, D. Sharpe, N. Jenkins, and E. Bossanyi: *Wind Energy Handbook*. John Wiley and Sons, Ltd (2001).
- [Castaing et al 1990] B. Castaing, Y. Gagne, and E. J. Hopfinger: Velocity probability density functions of high Reynolds number turbulence. In: *Physica D* 46, 177–200 (1990).
- [Friedrich et al 2000] R. Friedrich, S. Siegert, J. Peinke, St. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister: Extracting model equations from experimental data. In: *Phys. Lett. A* 271, 217–222 (2000).
- [Friedrich et al 2002] R. Friedrich, Ch. Renner, M. Siefert, and J. Peinke: Comment on “Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data”. In: *Phys. Rev. Lett.* 89, 149401 (2002).

- [Gardiner 1985] C. W. Gardiner: *Handbook of Stochastic Methods*. Springer, Berlin (1985).
- [Gottschall & Peinke 2007] J. Gottschall and J. Peinke: Stochastic modelling of a wind turbine's power output with special respect to turbulent dynamics. In: *J. Phys.: Conf. Ser.* 75, 012045 (2007).
- [Gottschall & Peinke 2008a] J. Gottschall and J. Peinke: How to improve the estimation of power curves for wind turbines. In: *Environ. Res. Lett.* 3, 015005 (2008).
- [Gottschall & Peinke 2008b] J. Gottschall and J. Peinke: Power curves for wind turbines – a dynamical approach. In: *Proc. of EWEC conference* (published online) (2008).
- [Haken 1983] H. Haken *Synergetics – An Introduction*. Springer, Berlin (1983).
- [Hau 2005] E. Hau: *Wind Turbines - Fundamentals, Technologies, Applications, Economics*. Springer, Berlin (2005).
- [Hölzer 2008] M. Hölzer: Stochastische Beschreibung des Schalenkreuzanemometers. Diplomarbeit (diploma thesis), Carl von Ossietzky Universität Oldenburg (2008).
- [IEC 2005a] International Standard IEC 61400-12-1. *Wind turbines – Part 12-1: Power performance measurements of electricity producing wind turbines*. First edition 2005-12.
- [IEC 2005b] International Standard IEC 61400-1. *Wind turbines – Part 1: Design Requirements*. Third edition 2005-08.
- [Kaiser et al 2007] K. Kaiser, W. Langreder, H. Hohlen, and J. Højstrup: Turbulence Correction for Power Curves. In: *Wind Energy – Proc. Euromech Colloquium* ed J. Peinke, P. Schaumann and St. Barth (Springer, Berlin), 28–31 (2007).
- [Kleinhans & Friedrich 2007] D. Kleinhans and R. Friedrich: Quantitative Estimation of Drift and Diffusion Functions from Time Series Data. In: *Wind Energy – Proc. Euromech Colloquium* ed J. Peinke, P. Schaumann and St. Barth (Springer, Berlin), 129–134 (2007).
- [Nawroth & Peinke 2006] A. P. Nawroth and J. Peinke: Multiscale reconstruction of time series. In: *Phys. Lett. A* 360, 234–237 (2006).
- [Panofsky & Dutton 1984] H. A. Panofsky and J. A. Dutton: *Atmospheric Turbulence*. John Wiley and Sons, Inc. (1984).
- [Ragwitz 2001] M. Ragwitz: Datengetriebene Modelle stochastischer dynamischer Systeme am Beispiel hydrodynamischer Turbulenz. Dissertation (PhD thesis), Bergische Universität Gesamthochschule Wuppertal (2001).
- [Rauh et al 2007] A. Rauh, E. Anahua, St. Barth, and J. Peinke: Phenomenological Response Theory to Predict Power Output. In: *Wind Energy – Proc. Euromech Colloquium* ed J. Peinke, P. Schaumann and St. Barth (Springer, Berlin), 153–158 (2007).
- [Risken 1989] H. Risken: *The Fokker-Planck Equation*. Springer, Berlin (1989).
- [Risø-M-2632] Risø-M-2632: *Accuracy of Power Curve Measurements*. C. J. Christensen and J. B. Dragt (Editors) (1986).

- [van der Hoven 1957] I. van der Hoven: Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. In: *J. Meteorol.* 14, 160–164 (1957).
- [Wächter et al 2008] M. Wächter, A. Rettenmeier, M. Kühn, and J. Peinke: Wind velocity measurements using a pulsed LIDAR system: first results. In: *IOP Conf. Ser.: Earth Environ. Sci.* 1, 012066 (2008).
- [Wagner et al 2008] R. Wagner, H. E. Jørgensen, U. S. Paulsen, T. J. Larsen, I. Antoniou, and L. Thesbjerg: Remote sensing used for power curves. In: *IOP Conf. Ser.: Earth Environ. Sci.* 1, 012059 (2008).

Chapter 4

Stochastic modelling of human postural control

In this chapter, it is shown by means of a further example of application how the Langevin approach can be utilized to model the dynamics of a set of experimental data. The reconstruction of drift and diffusion coefficients is applied to evaluate balance performance with respect to an experiment on a balance board. In particular, it is investigated to what extent supra-postural tasks influence the balance performance of respective test participants. A modification of the general reconstruction procedure is introduced that is particularly adapted to very short data sets, and it is shown how it can be evaluated if certain deviations in derived drift and diffusion coefficients are significant or not by means of an analysis of variances (ANOVA). The first section of this chapter is given by a publication that summarizes the general framework for the specific application. In the second section, a more advanced study is presented. In this second study, the differences between standing and sitting on the balance board have been investigated and, at this, it is distinguished between task and time effects on the respective balance performance.

4.1 Exploring the dynamics of balance data – movement variability in terms of drift and diffusion ¹

We introduce a method to analyze postural control on a balance board by reconstructing the underlying dynamics in terms of a Langevin model. Drift and diffusion coefficients are directly estimated from the data and fitted by a suitable parametrization. The governing parameters are utilized to evaluate balance performance and the impact of supra-postural tasks on it. We show that the proposed method of analysis gives not only self-consistent results but also provides a plausible model for the reconstruction of balance dynamics.

4.1.1 Introduction

Human movement coordination is essentially characterized by variability. Within studies on a wide range of experiments, including e.g. the research on gait dynamics, isometric force production or coordinated discrete and rhythmic movements, a couple of different measures have been introduced to quantify the observed variability in movement and evaluate a respective task performance [Newell & Corcos 1993, Davids et al 2005, Hausdorff 2007, Slifkin & Newell 1999].

The definition of an appropriate measure for a specific context is directly related to the question regarding the nature of the dynamical process that underlies the observed variability in human movement. Fluctuations in the variables that are measured in a respective experiment are in the simplest case due to uncorrelated noise and captured by summary statistics as mean values or standard deviations of the particular variables in an adequate way. More often, however, one can identify a dynamical process that induces finite correlations. In this case, a dynamical analysis can provide further information that can be used to characterize the observed process. Irregular fluctuations may originate from a nonlinear deterministic process or a random process as for example fractional Brownian motion. Respective methods for the analysis of human movement data utilizing advanced measures were presented e.g. in [Stergiou et al 2004] and [Collins & De Luca 1993].

In this work, we focus on an approach that combines both, deterministic and stochastic features. In particular, we present how the data of a balance experiment, as an example for investigating human movement control, can be analyzed in the framework of a Langevin process. For this class of stochastic processes, the interplay between deterministic and stochastic components is captured in terms of drift and diffusion. The drift part gives the deterministic behaviour of the considered system and can be best illustrated by a potential within that the movement takes place. The diffusion part adds dynamical noise that lets the movement fluctuate in this potential and that, in contrast to external measurement noise, directly influences the evolution of the dynamics. In studying turbulence, Friedrich and Peinke introduced a method how to quantify drift and diffusion for systems that exhibit the dynamics of a Markov process [Friedrich & Peinke 1997]. The method is based on well-known statistical concepts and can be applied directly to a measured time series. Over the years, this procedure has been utilized for the analysis of many different systems in a variety of disciplines including the analysis of tremor data, financial time series or heart-rate fluctuations [Friedrich et al 2000, Renner et al 2001, Kuusela 2004], as well as for the characterization of rhythmic human movement [Van Mourik 2006a].

¹Published as J. GOTTSCHALL, J. PEINKE, V. LIPPENS, and V. NAGEL: Exploring the dynamics of balance data – movement variability in terms of drift and diffusion. In: *Physics Letters A* 373, 811–816 (2009).

The different kinds of applications can be divided into the two categories of complexity in time and complexity in scale. Applications in the field of movement sciences normally belong to the first class and are described in terms of an autonomous Langevin equation. To simplify the specific reconstruction procedure, it is often required that the sampled signal corresponds to a stationary diffusion process – in general, the procedure is however not restricted to stationary processes (cf. e.g. [Van Mourik 2006b]). An explicit motivation of the Langevin approach is given in [Frank et al 2006]. The authors particularly motivate the use of the Langevin model for the analysis of human motor control variability by referring to the existence of different levels of description, a macroscopic and a microscopic, corresponding to deterministic and stochastic forces as stated in the synergetics approach (cf. [Haken 2004]).

By means of the experiment referred to in this contribution, we particularly investigate the performance of dynamical balance in combination with a supra-postural task. Similar studies were already reported in [Stoffregen et al 2000] and [Lippens 2005a]. Contrary to the prediction of the concept of resources (cf. [Woollacott & Shumway-Cook 2002, Fraizer & Mitra 2008]), it has been shown that the performance of balance does not decrease in the presence of supra-post tasks but rather increases in some cases. This result supports the hypothesis that postural control is not executed autonomously but is functionally integrated as a part of a broader information-movement system [Stoffregen et al 2007, Bootsma 1998]. In a number of studies, it was observed that the variability of postural sway is reduced to enhance perceptual contact to the environment in order to facilitate different supra-postural tasks (looking, reading resp. detecting [Stoffregen et al 1999, Stoffregen et al 2000, Stoffregen et al 2007], touching [Holden et al 1987, Riley et al 1999], or pointing [Balasubramaniam et al 2000]). The study presented in this contribution is directly attached to these investigations, considering however a specific measurement setup and a for this application novel method of data analysis.

4.1.2 Experimental setup and data analysis

A. Experimental setup

For the study presented in this contribution, we have considered a balance experiment and show, in particular, how the performance of postural control on a balance board can be analyzed in terms of drift and diffusion. A conventional balance board is made up of a hemisphere, the balance disc, and a circular platform on top of it on which the test participants stand and try to balance. Inside the hemisphere three gyros and three accelerometers measure the angular velocities ($\omega_\phi, \omega_\theta, \omega_\psi$) and accelerations (a_ϕ, a_θ, a_ψ) of the board. From the velocity and acceleration data the corresponding values for the angles (ϕ, θ, ψ) are reconstructed, for details see [Wagner et al 2003]. To minimize the dimensions of the state space and the needed number of measured data points, the degrees of freedom of the balance board were reduced for the present study by a mechanical construction, connecting in fact two balance discs by a prolate platform to a more rigid structure. Due to this modification, the movement of the balance board is restricted to only one direction in the horizontal plane spanned by ϕ and θ . For the present experiment, we considered the movement in medial-lateral direction corresponding to the angle ϕ (see figure 4.1).

The group of test participants consisted of 10 volunteering students (5 females, 5 males; age: 20–25 years). All of them were well experienced on the balance board. Each subject performed four trials on the balance board, two trials with supra-postural task and two without. The supra-postural task consisted in searching

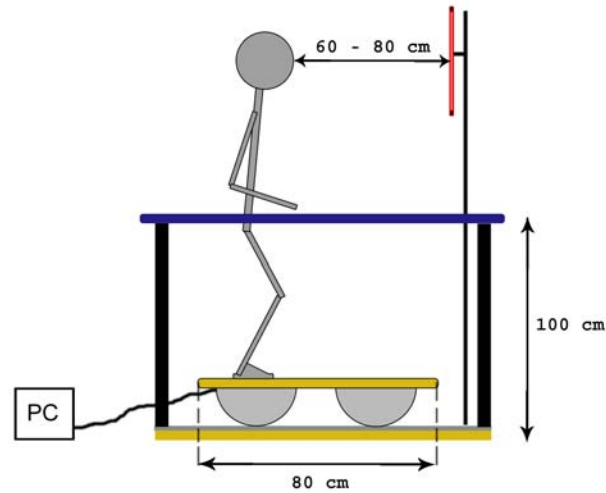


Figure 4.1: Experimental setup – consisting of the modified balance board (two balance discs connected by a platform on the top, reducing degrees of freedom to movement in only one direction), connection to a computer, surrounding handrail (to be used only before and after the data recording), and text on tripod for searching condition.

in one trial the letter E and for the other the letter H in a text displayed on a poster. The letters were presented in the same block of a continuous text but with different frequencies of occurrence (E: 102, H: 25). The participants counted the respective letters during the trial and reported the number of detected letters after the measurement was finished. In the trials without supra-postural task, the participants were asked to focus a marked point at the wall that was at the same position as the poster for the trials with searching. The subjects stood in bipedal stance on the board, and the sequence of trials was randomly arranged for each subject. For each trial, data was collected over a duration of 45 seconds and with a sampling rate of $f = 100$ Hz.

In the following, we firstly present how to model balance dynamics as a Langevin process and reconstruct its governing coefficients exemplarily for one data set. After that, we compare the results for the whole group of subjects and trials and discuss the effect of the supra-postural tasks on the balance performance.

B. Data analysis

Figure 4.2 shows an exemplary time series for the angle ϕ that describes the angular displacement in medial-lateral direction as well as for the corresponding angular velocity ω_ϕ . While $\phi(t)$ represents apparently a non-stationary process (or possibly the consequences of fluctuations of a longer stationary process, cf. [Duarte & Zatsiorski 2000], which may have for our approach the same relevance), $\omega_\phi(t)$ fluctuates around an equilibrium point of approximately zero velocity. Since our specific statistical approach requires the stationarity of the data, we applied the dynamical analysis to the data of angular velocities. The temporal variation of the angular velocity corresponds to a force or torque which is, furthermore, in a mechanical approach the appropriate quantity to describe the system. To simplify

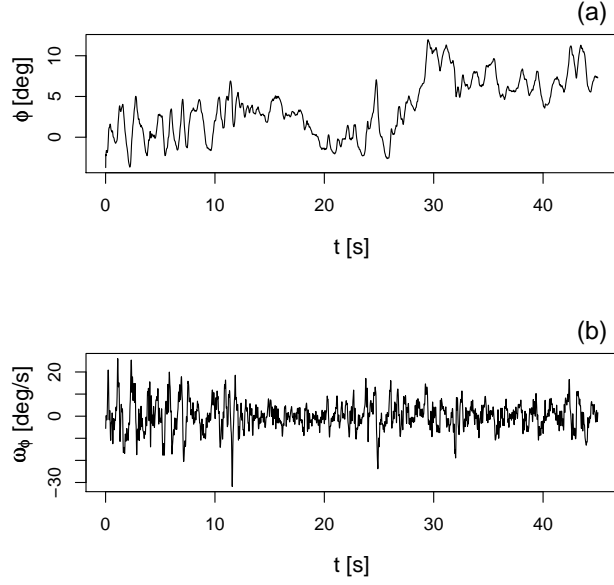


Figure 4.2: Exemplary time series for the angular displacement in medial-lateral direction and the corresponding angular velocity.

the notation we set $x(t) \equiv \omega_\phi(t)$.

We analyze the dynamics of $x(t)$ in terms of an autonomous Langevin process described by the first-order differential equation

$$\dot{x}(t) = D^{(1)}(x) + \sqrt{D^{(2)}(x)} \Gamma(t) . \quad (4.1)$$

The term $D^{(1)}(x)$ is called drift coefficient and reflects the deterministic part of the dynamics. Its stochastic counterpart is given by the Langevin force $\Gamma(t)$, representing Gaussian white noise with $\langle \Gamma(t) \rangle = 0$ and $\langle \Gamma(t_1) \Gamma(t_2) \rangle = 2\delta(t_1 - t_2)$ (following the convention in [Risken 1989]), and the square root of the diffusion coefficient $D^{(2)}(x)$, fixing the amplitude of the stochastic fluctuations. We apply Itô's interpretation of stochastic integrals here. $D^{(1)}(x)$ and $D^{(2)}(x)$ are explicitly not time-dependent.

In a first-order approximation for finite sampling rates, drift and diffusion coefficients are derived according to

$$M^{(n)}(x, \tau) = n! \tau D^{(n)}(x) + \mathcal{O}(\tau^2) \quad (4.2)$$

with the conditional moments

$$M^{(n)}(x, \tau) = \langle [x(t + \tau) - x(t)]^n \rangle \Big|_{x(t)=x}, \quad (4.3)$$

($n = 1, 2$). The time increment τ is given in units of the inverse sampling rate, i.e. f^{-1} . Note that the approximation $D^{(n)}(x) \approx M^{(n)}(x, \tau)/(n! \tau)$, as deduced from Eq. (4.2), is only exact for $\tau \rightarrow 0$. With increasing values of τ as well as in the presence of additional external noise, the estimated drift and diffusion coefficients systematically deviate from the intrinsic functions, and further correction terms must be taken into account for an exact reconstruction (cf. [Friedrich et al 2001, Böttcher et al 2006, Gottschall & Peinke 2008]). These deviations are not further

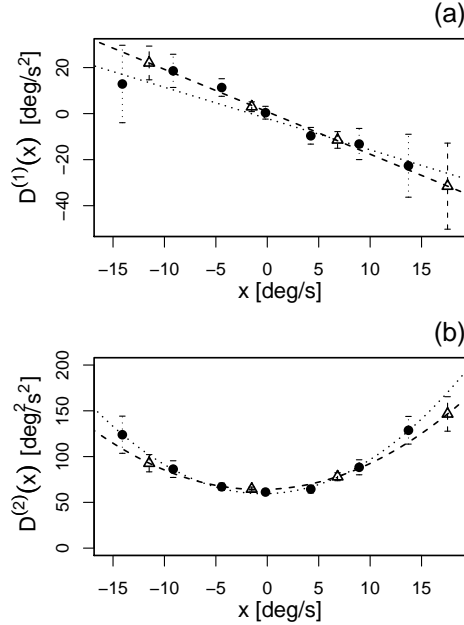


Figure 4.3: Reconstruction of drift and diffusion coefficient with the use of a binning (dashed lines and triangles for $N = 4$, dotted lines and dots for $N = 7$), and parametrizations applying a least-squares fit – (a) linear fit for $D^{(1)}(x)$ ($N = 4$: $a_0 = 0.80(\pm 0.17)$, $a_1 = -1.84(\pm 0.01)$; $N = 7$: $a_0 = -2.1(\pm 2.0)$, $a_1 = -1.36(\pm 0.17)$), and (b) second-order polynomial for $D^{(2)}(x)$ ($N = 4$: $b_0 = 64.2(\pm 0.2)$, $b_1 = 0.35(\pm 0.01)$, $b_2 = 0.25(\pm 0.01)$); $N = 7$: $b_0 = 59.5(\pm 1.1)$, $b_1 = 0.27(\pm 0.04)$, $b_2 = 0.34(\pm 0.01)$).

considered for the present analysis to simplify matters. It is assumed that their impact on the final results is negligible for this application. Though, a detailed analysis is beyond the scope of the present study.

In practice, the drift and diffusion coefficient are usually estimated by introducing a binning for the variable x , i.e. finding a discretization ($x_0, x_1, \dots, x_i, \dots, x_N$) of N equally sized intervals ($I_1, \dots, I_i, \dots, I_N$) and deriving the mean values

$$D^{(n)}[i] = \langle d_n(t) | x(t) \in I_i \rangle \quad (4.4)$$

with

$$d_n(t) = [x(t + \tau) - x(t)]^n / (n! \tau). \quad (4.5)$$

In detail, the boundaries of the bins are found by computing the minimum and the maximum of the sampled data and dividing the range spanned by these two values through the preliminarily defined number of bins. Additionally, a minimum number of data points per bin is specified which may decrease the actual number of bins. Figures 4.3(a) and (b) show the results for the data displayed in figure 4.2, applying two different binnings with $N = 4$ and $N = 7$ bins, respectively. The values for the drift coefficient were fitted by the linear function $y = a_0 + a_1 x$, the values for the diffusion coefficient by the second-order polynomial $y = b_0 + b_1 x + b_2 x^2$, using in each case the method of least squares. The error bars shown in the two figures are defined as standard errors for the values $D^{(1)}[i]$ and $D^{(2)}[i]$ in each bin. The uncertainties for the fitted parametrizations are given in the caption.

At first sight, the results of the parametrizations for the two different binnings seem to deviate from each other. Both results however exhibit quite large uncertainties that are dominated by the one or two outer points $D^{(n)}[i]$ ($n = 1, 2$),

respectively. Considering these uncertainties, the results are actually to a large extent consistent. Nevertheless, the plots illustrate that the introduction of a binning for data sets that consist only of a small number of data points involves certain difficulties. Due to the small total number of data points, we have to choose relatively large bin widths in order to keep a specified minimal number of data points per bin, which increases the uncertainty for the single points $D^{(n)}[i]$. Furthermore, single bins may be ignored when this minimal number is not reached. The single points for the bin averages $D^{(n)}[i]$ may give the impression of a larger precision than it is actually provided by the data basis.

An alternative procedure, avoiding primarily the errors induced by the binning, is to apply a least-squares fit directly to the values $d_n(t)$ versus $x(t)$. Respective results are shown in figure 4.4 and discussed in comparison to the former results in the next subsection.

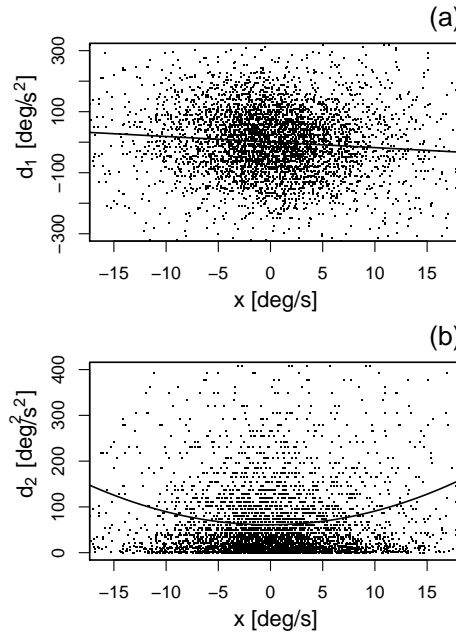


Figure 4.4: Direct reconstruction of drift and diffusion coefficient without binning, and parametrizations applying a least-squares fit – (a) linear fit for $D^{(1)}(x)$ ($a_0 = 0.49(\pm 1.79)$, $a_1 = -1.82(\pm 2.8)$), and (b) second-order polynomial for $D^{(2)}(x)$ ($b_0 = 60.9(\pm 2.3)$, $b_1 = -0.07(\pm 0.32)$, $b_2 = 0.29(\pm 0.03)$).

C. Self-consistency test

To check the quality of the results of both procedures, we performed a self-consistency test. Therefore, we estimated the empirical probability density function (pdf) directly from the data and compared it with a theoretical pdf that is derived on the basis of the reconstructed $D^{(1)}(x)$ and $D^{(2)}(x)$. Following [Risken 1989], the stationary pdf of a Langevin process defined by (4.1) is given by

$$p_{stat}(x) = \frac{\mathcal{N}}{D^{(2)}(x)} \exp\left(\int^x d\tilde{x} \frac{D^{(1)}(\tilde{x})}{D^{(2)}(\tilde{x})}\right) \quad (4.6)$$

where \mathcal{N} is a normalization constant. The results for the exemplary data set of figure 4.2 and the corresponding estimates for drift and diffusion coefficient in figures 4.3

and 4.4 are shown in figure 4.5. The theoretical pdfs are based on the obtained parametrizations, i.e. the fitted functions for drift and diffusion, and Eq. (4.6). The different results for the method of reconstruction with binning, dependent on how many bins were used, recur in two different results for the derived pdf (dashed line for $N = 4$, dotted line for $N = 7$). A third theoretical pdf is given by the results of the direct reconstruction procedure. The empirical pdf is determined by the histogram of the measured data (given by symbols in a semi-logarithmic plot). Note that the discretization applied at this has a much finer resolution than the binning used for the estimation of drift and diffusion coefficient – the outer bins include only a few data points. (The value of $p(x) = 4 \cdot 10^{-4}$, which is the lowest probability displayed, corresponds to only two data points.) The goodness of fit, i.e. the agreement between the empirical and the reconstructed pdfs, $p(x)$ and p_{ref} , respectively, depends strongly on the chosen binning.

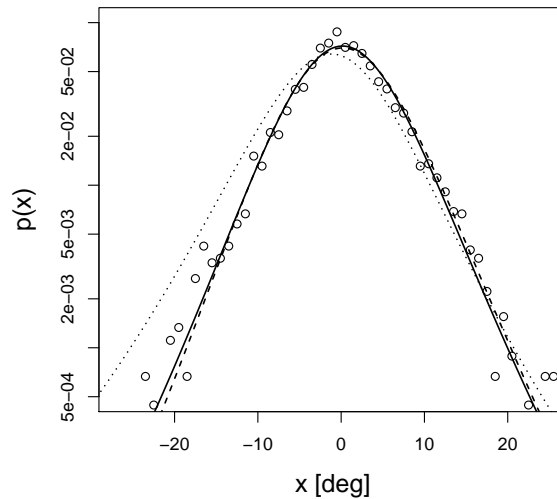


Figure 4.5: Empirical pdf (histogram given by dots) and corresponding functions reconstructed according to Eq. (4.6) applying the methods with (dashed and dotted lines for $N = 4$ and $N = 7$ bins, respectively) and without binning (solid line).

Deriving the respective Chi-square distances

$$d_C(p(x), p_{\text{ref}}(x)) = \frac{\int_{-\infty}^{+\infty} dx [p(x) - p_{\text{ref}}(x)]^2}{\int_{-\infty}^{+\infty} dx p_{\text{ref}}(x)}, \quad (4.7)$$

replacing the integrals by finite summations for a finite number of points, we find the values 0.034 and 0.131 (for $N = 4$ and $N = 7$, respectively) for the method with binning. Applying the direct method, we obtain a Chi-square distance of 0.031 between the reconstructed pdf and the empirical pdf. Evaluating a respective Chi-square test, i.e. comparing the distributions in terms of a conventional Chi-square statistics and defining thereby significance e.g. on a level of 5% , we find that the hypothesis of an agreement with the empirical pdf must be rejected for all three reconstructed pdfs. This result is not surprising since the empirical pdf is based on only a small data basis and is, furthermore, essentially affected by

the used discretization – the reconstructed pdfs are however smoothed by means of the parametrization. Note also that the single data points are not statistically independent but governed by a deterministic relation. The Chi-square measure is hence only a relative measure. Nonetheless, it may serve as a criterion to evaluate the reconstructed results with respect to their consistency with the empirical data. In this regard, the found values for $d_C(p(x), p_{\text{ref}}(x))$ may argue for the direct method in comparison to the method with binning. We obtained similar outcomes also for the other data sets, and could confirm them with the analysis of typical numerical data.

A simple explanation for this result is that we consider for the estimation of drift and diffusion with binning not all of the measured data points but include only the points in those bins containing a certain minimum of data. Removing some data induces a kind of asymmetry in comparison to the determined empirical pdf that contains all the data points. The advantage of the direct method, in contrast, is that the complete data set is considered and no data limitations must be defined.

Consequently, we can conclude that a direct method of reconstruction without binning should be especially for short data sets the method of choice. It is however necessary to mention that the direct method is only applicable if drift and diffusion coefficient are in a simple relation to the conditional moments as given in Eq. (4.2). For lower sampling rates or in the presence of external measurement noise the relation is more complex (cf. [Gottschall & Peinke 2008]) and a binning is necessary for the analysis. Furthermore, the direct method requires the previous knowledge of the functional forms of $D^{(1)}(x)$ and $D^{(2)}(x)$, so that it is recommendable to perform a pre-analysis using the method with binning.

4.1.3 Results

A. Performance-of-balance measures

To examine the influence of supra-postural tasks on postural control, and in detail of visual searching on the performance of postural control on the balance board, we analyzed the data for each subject and each trial recorded for the study described above in the following way. The first step has been to define an appropriate measure expressing the degree of movement variability. In respect of the analysis in terms of drift and diffusion, presented in the preceding section, we utilized the single reconstructed parameters of $D^{(1)}(x)$ and $D^{(2)}(x)$ as a set of independent performance measures. In doing so, we presume that a change in movement variability is given by a quantitative change in at least one of the listed parameters. (In the following, we refer to this approach as *method B*.) This we compare with the approach in [Lippens 2005b], where the degree of variability and the corresponding performance of balance is measured in terms of the root mean square (RMS) values for all three degrees of freedom as a sample statistic for the single data sets. As a slight modification, we consider in this paper the standard deviations (SD) for the respective time series of angular velocity in medial-lateral direction (*method A*).

Both approaches, *method A* and *B*, are linked according to Eq. (4.6). The reconstructed drift and diffusion coefficients can be utilized to split up the standard deviation into a deterministic and a stochastic part. Moreover, they give not only the width of the pdf but its actual shape, and thus contain considerably more information than the single value SD.

B. Revealing significant effects by application of ANOVA

Having once defined the performance measure, we could apply an analysis of variances (ANOVA) with the performance measure as dependent variable to investigate

Table 4.1: Overview of main effects revealed by an one-way ANOVA on the factor task ($k = 2$ and $k = 4$ levels, resp.) for different (partial) measures. Significant effects are indicated by * ($p < 0.05$).

Variable	F(1, 38)	p<	F(3, 36)	p<
<i>method A:</i>				
standard deviation (SD)	15.778	0.000 *	5.226	0.004 *
<i>method B:</i>				
a_0 (const. drift part)	0.584	0.449	0.201	0.895
a_1 (linear drift part)	0.841	0.373	0.349	0.790
b_0 (const. diffusion part)	10.340	0.003 *	4.005	0.015 *
b_1 (linear diffusion part)	2.514	0.121	1.176	0.332
b_2 (quadratic diffusion part)	0.845	0.364	2.199	0.105

possible systematic task-related differences in the degree of variability for the whole group of test participants. We performed a one-way ANOVA on the factor task with $k = 2$ and $k = 4$ factor levels, respectively. (For details about this kind of statistical test we refer e.g. to [Chambers et al 1992].) For the first part of the analysis, we distinguished solely between the trials with and the trials without supra-postural task, resulting in two treatments or levels (searching vs. no searching). Originally, the study was designed to be composed of four treatments, corresponding to the four trials for each participant (first and second trial without task, searching the letters E and H). Since the two trials without supra-postural task differ only with respect to their order, we however have intended to investigate task effects but not the impact of time or learning with this first approach, an analysis with four levels has to be considered with caution. We present the results of both ANOVA designs here. For the respective discussion, we have defined significance on a level of $p < 0.05$.

Evaluating the values for SD as dependent variable (*method A*), we found in accordance with [Stoffregen et al 2007] a significant task effect, expressed by a decreased standard deviation of angular velocity in medial-lateral direction during searching that corresponds to an increase of balance performance, for both the ANOVA with two and with four treatments ($k = 2$: $F(1, 38) = 15.778$, $p < 0.0003$; $k = 4$: $F(3, 36) = 5.226$, $p < 0.004$). To classify the task effect with respect to the ANOVA with four factor levels further, we applied a Tukey HSD test that compares all possible pairs of means for the different factor levels analyzed for the ANOVA and identifies on the basis of a studentized range distribution the significant differences. Here, the decomposition of the factor task shows a significant increase of balance performance for searching conditions compared with the first no-task condition (E: $p < 0.018$, H: $p < 0.019$). Compared to the second no-task condition the effect is not significant (E: $p < 0.094$, H: $p < 0.099$).

We performed the same analysis for the single parameters of $D^{(1)}(x)$ and $D^{(2)}(x)$ applying a linear and second-order parametrization, respectively, as given above (*method B*). For both applied designs, the ANOVA solely reveals a main effect of task for the parameter b_0 , i.e. the constant part of the diffusion coefficient ($k = 2$: $F(1, 38) = 10.340$, $p < 0.0027$ and $k = 4$: $F(3, 36) = 4.005$, $p < 0.015$; cf. figure 4.6). All other parameters show no significant effects – see Tab. 1. Applying a Tukey HSD test, a decomposition of the factor task for the design with four factor levels again shows a significant increase in performance of balance for the searching conditions compared to the first no-task condition (E: $p < 0.030$, H: $p < 0.028$) but no significant effect in comparison to the second no-task condition (E: $p < 0.398$,

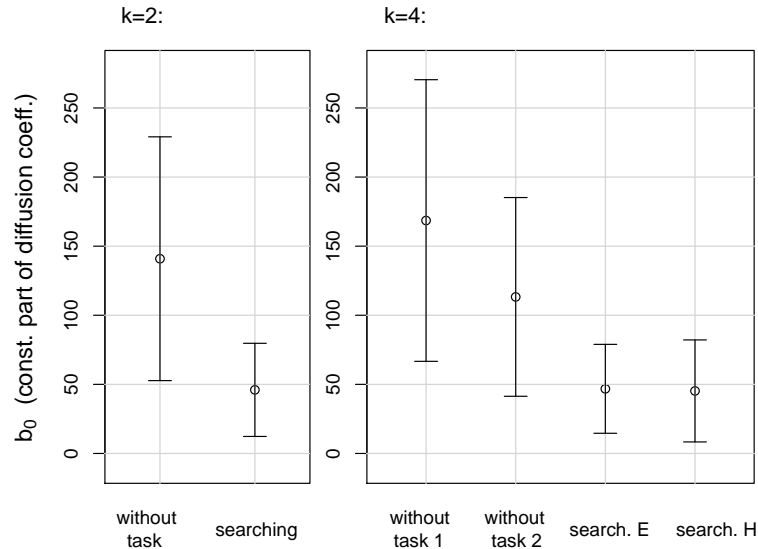


Figure 4.6: Performance of balance with and without supra-postural tasks in terms of the constant part of the diffusion coefficient. Shown are the mean values of the measure for each factor level (ANOVA with $k = 2$ and $k = 4$ treatments, resp.), vertical bars indicate a 0.95-confidence interval. The performed ANOVAs reveal significant effects with $F(1, 38) = 10.340$, $p < 0.003$ and $F(3, 36) = 4.005$, $p < .015$, resp.

H: $p < 0.378$).

Beyond, an ANOVA with the factor time (1st–4th trial) gives no significant effects for all used performance measures, so that we can rule out the possibility of a general impact of the arrangement of the trials.

Hence, the analysis of drift and diffusion confirms the former results based on the comparison of standard deviations and provides further more detailed information about the process dynamics beyond it. A decrease in the standard deviation of the angular velocity for balancing with supra-posural task in comparison to balancing without additional task is thus primarily traced back to a difference in the stochastic part of the dynamics but not in the deterministic part. That is, in the presence of a supra-postural task, balance is performed with less dynamical noise in the corresponding Langevin process. In principle, a reduced variability, given by a smaller value for the standard deviation of the angular velocity, could also be achieved by an increased slope of $D^{(1)}(x)$, a larger absolute value for a_1 , i.e. a faster deterministic reaction. This is however not observed here.

4.1.4 Conclusions

To summarize, we have introduced a dynamical method for the analysis of balance data based on the reconstruction of an underlying Langevin process. For this purpose, we have applied a well established procedure but particularly adapted it to

very short data sets. Our analysis confirms former results with regard to the impact of supra-postural tasks on the balance performance, but provides further insights into the process dynamics separating deterministic and stochastic parts. In a further study we are going to show how this separation can be utilized to differentiate between task effects and effects of training in repeated balance experiments. For the present data, we showed the self-consistency of our results, and approved the initial assumption that the process can be described by a Langevin process. To estimate drift and diffusion coefficients from the measured data, we applied a first-order approximation for the conditional moments. Systematic deviations that may arise due to too low sampling rates or in the presence of external measurement noise and that would possibly require the consideration of higher-order terms were not discussed in detail. It is assumed that respective systematic errors do not influence the final results of the experiment, i.e. the outcome of the performed ANOVA, since they affect the single trials respectively in the same manner. Altogether, our findings suggest to apply the described approach to analyze human motor control variability and reconstruct human movement coordination in terms of drift and diffusion.

4.2 Distinguishing between task and time effects – a study on rowers standing and sitting on a balance board ²

4.2.1 Outline of the study

Objective of the study, that is discussed in this section, has been to evaluate the balance performance of rowers in a standing and a sitting condition, and particularly analyze the influence of supra-postural tasks as well as certain time effects.

The group of test participants has consisted of in total 13 junior rowers of a regional rowing club in Hamburg (mean age: 13.14 years), though not all participants attended each measurement. The study was carried out during winter and spring 2008 and has been divided into five measurements ($t_I - t_V$) under varying conditions with N participants in each case – two pre-test measurements (t_I : standing 1, $N=6$; t_{II} : sitting 1, $N=7$) during the winter training, two post-test measurements after a period of approximately two months (t_{III} : sitting 2, $N=6$; t_{IV} : standing 2, $N=5$) and lastly a retention measurement after again two months (t_V : sitting 3, $N=7$).

As measurement instrument, the modified version of the balance board was used (see 4.1.2). The subjects either stood in bipedal stance (cf. figure ??) or sat, with their feet on an additional balance disc, on the board – the measurement setup for the sitting condition is illustrated in figure 4.7.

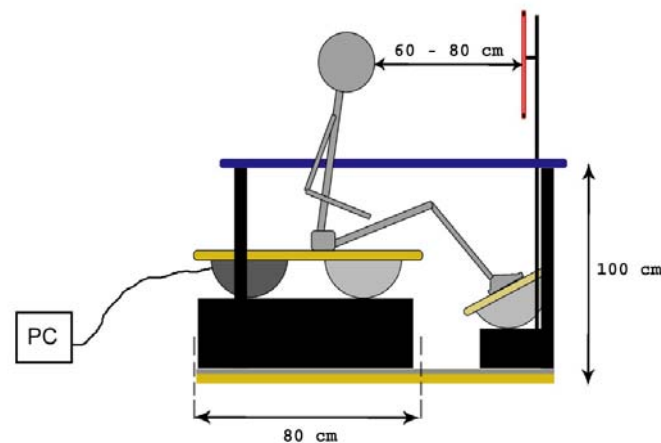


Figure 4.7: Experimental setup for sitting condition – consisting of balance board with connection to a computer, surrounding handrail, and poster-text on tripod for searching condition.

For each measurement, the individual participants underwent four single but successive trials on the balance board – each with a duration of 45 sec. Thereof, they performed two trials with a supra-postural task (searching the letters E or H) and the two other without – just as described in 4.1.2. The sequence of trials was systematically permuted across the participants, resulting in a random order for the individual subject. The data of the balance board was recorded with a sampling rate of $f = 100$ Hz.

²The results described in this section were presented as V. LIPPENS, V. NAGEL, J. GOTTSCHALL, and J. PEINKE: Performance of balance – General motor ability or specific adaption of strategies? at the WCPAS VIII conference (2008), and published under the same title in *World Congress of performance Analysis of Sport VIII. Book of Proceedings* eds A. Hökelmann and M. Brummund; Otto-von-Guericke University, Magdeburg, 354–359.

For the standing as well as for the sitting conditions, the velocity data of the movement in medial-lateral direction was analyzed. We accomplished a data reduction by deriving the standard deviations (SD) of the single data sets, i.e. for each trial of each participant and measurement, as well as the set of parameters $(a_0, a_1, b_0, b_1, b_2)$ applying the Langevin approach and a respective parametrization of the drift and diffusion coefficients, as it is exemplified in the last preceding section. The resulting six parameters were one after the other investigated by means of an ANOVA for repeated measures assuming different models. The explicit proceeding is described in the following section.

4.2.2 Results and discussion

The performance of balance of the rowers was analyzed in three different models – a trial model to investigate the impact of the order of trials (1st–4th trial for both groups of measurements, under standing and sitting conditions, respectively), a task model to evaluate the impact of the different kinds of supra-postural tasks (trial without task 1 and 2, searching the letter E, searching the letter H – likewise for both groups of measurements), and a time model to detect possible systematic differences between the single measurements (standing 1-2 and sitting 1-3, without further specifying the factor task or trial). For all three models, we performed an ANOVA with repeated measurements (cf. Appendix) for each single parameter – SD, and a_0, a_1, b_0, b_1, b_2 for the Langevin approach – as dependent variable, respectively. Thereby, significance is again defined on a level of $p < 0.05$.

Neither for the sitting nor for the standing condition, we could find an effect of trial. I.e. the balance performance of a particular trial does not depend on the position within the series of trials at that it is performed. An effect of task could be found for the standing conditions (t_I, t_{IV}), and that for several parameters (for SD a significant effect with $F(3, 27) = 5.446$, $p < 0.005$, and for b_0 a weak tendency given by $F(3, 27) = 2.512$, $p < 0.080$), but not for the sitting conditions of the pre- and post-test measurements (t_{II}, t_{III}). A task effect for a sitting condition could however be observed for the retention measurement (t_V) – again for the parameters SD and b_0 ($F(3, 18) = 6.181$, $p < 0.005$ and $F(3, 18) = 5.251$, $p < 0.020$, respectively). Significant time effects, in contrast, were found only for the sitting conditions (t_{II}, t_{III}, t_V) but not for the standing conditions (t_I, t_{IV}). For the former case, highly significant effects were detected for the parameters SD ($F(2, 17) = 13.824$, $p < 0.000$), b_0 ($F(2, 17) = 7.977$, $p < 0.004$), and in addition for a_0 ($F(2, 17) = 59.059$, $p < 0.000$).

The results for the measurements under standing conditions are consistent with former studies, and especially with the conclusions of 4.1, confirming the idea of a functional integration of postural control (see e.g. [Stoffregen et al 2007]). The outcome of the Langevin approach confirms the results obtained on the basis of the standard deviation of the angular velocity data. For the most cases, a decrease in SD is directly related to a smaller value for b_0 , i.e. the constant part of the diffusion coefficient in the Langevin model. Under the sitting conditions, we conclude, the rowers have to obtain a certain level of performance before the supra-postural tasks facilitate the motor action (cf. [Bloem et al 2006]). This may explain why a task effect is not observed until the retention measurement. The performance of balance in the sitting condition is significantly increased during the line of the pre-test, the post-test and the retention measurement – for the standing conditions this development is not observed. One may conclude that, comparing standing and sitting on the balance board, the balance control of rowers is caused by different adaptations of strategies. As last point, we want to comment the observed significant time effect for the parameter a_0 in the sitting condition. This effect can be explained with regard to the measurement setup. Repeating the experiment at another time,

the test participants do usually not find exactly the same position they had for the first measurement on the balance board. Thus, the movement fluctuates around a slightly different, possibly less stable point. Accordingly, also the angular velocity may fluctuate around different fixed points that are not necessarily zero, resulting in an asymmetric and unstable or eventually non-stationary movement. Since the effect is observed for the whole group and not only for single participants, it can clearly be related to the setup. We conclude that a main effect for a_0 serves as a kind of indicator for a modified setup and is helpful for the interpretation of time effects.

4.3 Appendix: One-way ANOVA

For those readers who are not familiar to the application of an analysis of variance (ANOVA) we give here a short outline of the respective procedure as it is applied in sections 4.1 and 4.2. We restrict our considerations to a one-way ANOVA, i.e. we consider one independent variable or factor affecting one dependent variable. For more details and a comprehensive overview we e.g. refer to [Chambers et al 1992]).

For our applications, the dependent variable is the type of trial and the dependent variable is the respective performance measure. The values of the dependent variable are grouped with respect to different factor levels. The different levels are for our application either the specifications with or without supra-postural task (in detail, searching the letter E or H and first or second trial without task) or the number with respect to the order of trials (1–4th trial).

According to this grouping, one calculates the so-called variance of the group means

$$\text{Var}_1 := \sum_{j=1}^k n_j (\bar{X}_j - \bar{X})^2 / (k - 1) \quad (4.8)$$

as well as the mean of the within-group variances

$$\text{Var}_2 := \sum_{j=1}^k (n_j - 1) \sigma_j^2 / (N - k), \quad (4.9)$$

where k is the number of groups or factor levels, respectively, n_j the number of trials for each factor level and N the total number of trials. \bar{X}_j and σ_j^2 denote the mean value and the variance of the dependent variable within each group j , and \bar{X} is the mean value of all N trials. The quotient

$$F := \text{Var}_1 / \text{Var}_2 \quad (4.10)$$

gives a test statistic that is evaluated by means of an F-test. That is, to test the null hypothesis that the mean values for the dependent variable of the different groups are equal, the value F is related to a Fisher's $F(k - 1, n - k)$ -distribution where the specifications in the parentheses give the degrees of freedom. The result is the probability of error with that the null hypothesis can be declined. For the discussion of the obtained results, significance is defined on a certain level as e.g. $p < 0.05$ – i.e. p-values smaller than this bound serve as indication for a significant difference in the dependent variable grouped according to the factor levels.

For the application in section 4.2, we performed instead of the standard one-way ANOVA, outlined above, a so-called one-way ANOVA for repeated measures. At this, the term repeated measures means evaluating multiple measures per subject, which refers for our application to the different measurements performed at different times. A repeated measures design increases the sensitivity of the test since the test subjects serve as their own controls, eliminating the problem of across-subject variations.

Bibliography

- [Balasubramaniam et al 2000] R. Balasubramaniam, M. A. Riley, and M. T. Turvey: Specificity of postural sway to the demands of a precision task. In: *Gait and Posture* 11(1), 12–24 (2000).
- [Bloem et al 2006] B. R. Bloem, Y. A. M. Grimbergen, J. G. van Dijk, and M. Munneke: The “posture second” strategy – A review of wrong priorities in Parkinson’s disease. In: *Journal of Neurological Sciences* 248, 196–204 (2006).
- [Böttcher et al 2006] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase: Reconstruction of Complex Dynamical Systems Affected by Strong Measurement Noise. In: *Phys. Rev. Lett.* 97, 090603 (2006).
- [Bootsma 1998] R. J. Bootsma: Ecological movement principles and how much information matters. In: *Models in Human Movement Sciences – Proc. of the 2nd international symposium of the institute for fundamental and clinical human movement science* eds A. A. Post, J. P. Pijpers, P. Bosch, and M. S. J. Boschker; Print Partners Ipskamp, Enschede (1998).
- [Chambers et al 1992] J. M. Chambers, A. Freeny, and R. M. Heiberger: *Statistical Models in Seds* J. M. Chambers and T. J. Hastie; Wadsworth and Brooks/Cole (1992).
- [Collins & De Luca 1993] J. J. Collins and C. J. De Luca: Open-loop and closed-loop control of posture – a random-walk analysis of center-of-pressure trajectories. In: *Exp. Brain Res.* 95(2), 308–318 (1993).
- [Davids et al 2005] K. Davids, S. Bennett, and K. M. Newell: *Movement System variability*. Champaign, III.: Human Kinetics (2005).
- [Duarte & Zatsiorski 2000] M. Duarte and V. M. Zatsiorsky: On the fractal properties of human standing. In: *Neurosc. Lett.* 283(3), 173–176 (2000).
- [Fraizer & Mitra 2008] E. V. Fraizer and S. Mitra: Methodological and interpretive issues in posture-cognition dual-tasking in upright stance. In: *Gait and Posture* 27, 271–279 (2008).
- [Frank et al 2006] T. D. Frank, R. Friedrich, and P. J. Beek: Stochastic order parameter equation of isometric force production revealed by drift-diffusion estimates. In: *Phys. Rev. E* 74, 051905 (2006).
- [Friedrich & Peinke 1997] R. Friedrich and J. Peinke: Description of a Turbulent Cascade by a Fokker-Planck Equation. In: *Phys. Rev. Lett.* 78, 863–866 (1997).
- [Friedrich et al 2000] R. Friedrich, S. Siegert, J. Peinke, St. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister: Extracting model equations from experimental data. In: *Phys. Lett. A* 271, 217–222 (2000).
- [Friedrich et al 2001] R. Friedrich, Ch. Renner, M. Siefert, and J. Peinke: Comment on “Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data”. In: *Phys. Rev. Lett.* 89, 149401 (2001).
- [Gottschall & Peinke 2008] J. Gottschall and J. Peinke: On the definition and handling of different drift and diffusion estimates. In: *New Journal of Physics* 10, 083034 (2008).
- [Haken 2004] H. Haken, *Synergetics: Introduction and Advanced Topics*. Springer, Berlin (2004).

- [Hausdorff 2007] J. M. Hausdorff: Gait dynamics, fractals and falls – Finding meaning in the stride-to-stride fluctuations of human walking. In: *Human Movement Science* 26(4), 555-589 (2007).
- [Holden et al 1987] M. Holden, J. Ventura, and J. R. Lackner: Influence of light touch from the hand on postural sway. In: *Society for Neuroscience Abstracts* 13 (348), 292–296 (1987).
- [Kuusela 2004] T. Kuusela: Stochastic heart-rate model can reveal pathologic cardiac dynamics. In: *Phys. Rev. E* 69, 031916 (2004).
- [Lippens 2005b] V. Lippens: In for a Penny, in for a Pound! Do we use central resources or a functional integration while balancing? In: *Proc. of European Workshop On Movement Science EWOMS 2005. Mechanics, Physiology, Psychology*. Köln, Strauß(2005).
- [Lippens 2005a] V. Lippens: Reading can facilitate balancing on the gyro. In: *Recherches Actuelles en Science du Sport. Actes du 11th Congrès International de l'ACAPS* eds N. Benguigui et al.; EDP Sciences, Monts (2005).
- [Newell & Corcos 1993] K. M. Newell, D. M. Corcos: *Variability and motor control*. Champaign, III.: Human Kinetics (1993).
- [Renner et al 2001] Ch. Renner, J. Peinke, and R. Friedrich: Evidence of Markov properties of high frequency exchange rate data. In: *Physica A* 298, 499–520 (2001).
- [Riley et al 1999] M. A. Riley, T. A. Stoffregen, M. J. Grocki, and M. Turvey: Postural stabilization for the control of touching. In: *Human Movement Science* 18, 795–817 (1999).
- [Riskin 1989] H. Riskin, *The Fokker-Planck Equation*. Springer, New York (1989).
- [Slifkin & Newell 1999] A. B. Slifkin and K. M. Newell: Noise, information transmission and force variability. In: *J. Exp. Psychol. Hum. Percept. Perform.* 25, 837–851 (1999).
- [Stergiou et al 2004] N. Stergiou, U. H. Buzzi, M. J. Kurz, and J. Heidel: Nonlinear tools in human movement. In: *Innovative analyses of human movement*. Champaign, III.: Human Kinetics, 63–90 (2004).
- [Stoffregen et al 1999] T. A. Stoffregen, L. J. Smart, B.G. Bardy, and R. J. Pagulayan: Postural Stabilization of Looking. In: *Journal of Experimental Psychology: Human Perception and Performance* 25 (6), 1641–1658 (1999).
- [Stoffregen et al 2000] T. A. Stoffregen, R. J. Pagulayan, B. G. Bardy, and L. J. Hettinger: Modulating postural control to facilitate visual performance. In: *Human Movement Science* 19, 203–220 (2000).
- [Stoffregen et al 2007] T. A. Stoffregen, P. Hove, B. G. Bardy, M. Riley, C. T. Bonnet: Postural stabilization of perceptual but not cognitive performance. In: *Journal of Motor Behavior* 39, 126–138 (2007).
- [Van Mourik 2006a] A. M. van Mourik, A. Daffertshofer, and P. J. Beek: Deterministic and stochastic features of rhythmic human movement. In: *Biological Cybernetics* 94, 233–244 (2006).
- [Van Mourik 2006b] A. M. van Mourik, A. Daffertshofer, and P. J. Beek: Extracting Kramers-Moyal coefficients from short data sets and non-autonomous systems. In: *Phys. Lett. A* 351, 13–17 (2006).

- [Wagner et al 2003] J. F. Wagner, V. Lippens, V. Nagel, M. Morlock, and M. Vollmer: An Instrument Quantifying Human Balance Skills – Attitude Reference System for an Ankle Exercise Board. In: *IJCSS (International Journal of Computer Science in Sport)*, Special Edition 1, 96–105 (2003).
- [Woollacott & Shumway-Cook 2002] M. Woollacott and A. Shumway-Cook: Attention and the control of posture and gait – a review of an emerging area of research. In: *Gait and Posture* 16, 1–14 (2002).

Chapter 5

Summary

The basic intention of this thesis has been to show the applicability of a particular method of stochastic modelling for complex systems, based on the reconstruction of effective Langevin equations, and to extend and improve the respective procedure of data analysis with respect to specific examples of application. Thereby, the method has been particularly extended to Langevin-like processes that are not determined by Langevin equations in a strict sense but for that a reconstructing based on the Langevin approach nevertheless provides useful information.

Subsequent to an overview about previous fields of application that were published during the last years, two specific novel applications were described in more detail in this thesis – the modelling of a wind turbine’s power performance in Chapter 3, and the investigation of human postural control in Chapter 4. With respect to the large variety of very different applications, represented also by these two examples, it suggests itself to ask questions like how universal the presented and applied method actually is, how good and significant the results for a specific application, and eventually which benefit the method provides in comparison to other, possibly more conventional methods. To give in some respect an answer to these questions, I have elaborated on different aspects that may help to evaluate a specific application with respect to the theoretical Langevin model that is assumed.

In Chapter 2, the influence of low sampling rates and the presence of additional external measurement noise on the results of the reconstruction procedure has been analyzed in detail. These two kinds of disturbances have been chosen because it is assumed that they are the most prominent ones for the application of the reconstruction procedure to experimental data and that they possibly have been ignored for several applications so far having led to inaccurate results consequently – which has already been annotated and discussed in former publications (cf. Chapter 2). In particular, it has been pointed out that a reconstruction for a process that is exposed to at least one of the two discussed disturbances may indicate a multiplicative noise term although the intrinsic process is defined by additive noise. This may lead to a significant misinterpretation of the stochastic part of the observed and modelled dynamics. Based on a detailed analysis of the impact of finite-resolution and measurement-noise effects on the reconstruction of the underlying stochastic processes for numerical data, the relevance of studying the behaviour of the conditional moments of the data as function of a respective time increment has been underlined. Corresponding deviations from a linear relation as well as non-zero offsets may serve as significant indicators for a specific kind of disturbance. These indicators may give a hint regarding the accuracy of the reconstructed drift and diffusion functions.

In the same context, also the robustness of reconstructed fixed points as characteristic steady states for the system under consideration has been studied. The

estimation of the fixed points of a dynamical system corresponds to a simple and especially, as exemplified in Chapter 3, very flexible approach to characterize a complex system. The above-mentioned disturbances influence the results for the fixed points much less than the overall behaviour of the respective drift and diffusion functions but their impact does not totally vanish. As a qualitative indicator, the asymmetries of the studied process were determined, with respect to those the accordant inaccuracy of the reconstructed fixed points may be evaluated.

In Chapter 3, I illustrated with the example of the reconstruction of a wind turbine's power performance how the concept of stochastic modelling may be interpreted as a procedure of an effective modelling. The interplay of the wind turbine with the turbulent wind fluctuations, where the latter are moreover measured in a certain distance to the turbine, corresponds to a typical complex system and motivates in this way the usage of a stochastic approach. The proposed stochastic relaxation model, given by a one-dimensional Langevin equation for the power output with respect to each single interval of fixed wind speed, describes the transfer of turbulence from one variable to the other. This was, in particular, illustrated by a detailed characterization of turbulent structures in measured wind speed and power output time series. With regard to the modelling procedure, it has been pointed out that the Langevin model describes the analyzed power performance dynamics only in a first approximation. The proposed set of Langevin equations is used as a practical tool to characterize the considered system but it is obvious from the analysis that it is not sufficient to reconstruct the actual process dynamics. At this, in particular, the problem of non-stationarity is overcome by a strongly simplified approach. As another aspect, the Langevin modelling as a tool with a relative meaning but without a total absolute significance involves certain definitions that primarily facilitate a practical application but that also incorporate a kind of arbitrariness.

In Chapter 4, dealing with the modelling of human postural control, it has been focused with the analysis of very short data sets on a further aspect. The results of the reconstruction procedure for the single data sets, i.e. the single trials in the experiment on the balance board, come along with a quite large uncertainty due to the small amount of data. To facilitate a reliable evaluation, the individual results are combined to different groups that are analyzed by means of an analysis of variances (ANOVA). Therewith, a framework for the determination of significance is defined that primarily forms a basis for utilizing and interpreting the results of the reconstruction. This procedure is generally recommended for applications that involve only short time series of data.

With respect to the methodology, all the presented investigations are based on, the aspects that have been discussed in this thesis can be summarized as follows. It has been distinguished between the intrinsic dynamics, defined by a Langevin or Langevin-like process, and the observed dynamics that may additionally include external noise superimposed to the intrinsic dynamics. The respective data that is sampled as a corresponding discrete time series is basically characterized by a finite sampling rate and a finite amount of data – restrictions that also may be interpreted as certain disturbances. An appropriate evaluation of the applicability of the presented Langevin approach thus consists in investigating how the intrinsic process dynamics is affected by the process of observation and data sampling, respectively.

A last remark refers to the aspect of variability with that the considerations in the introduction in Chapter 1 were opened. It has been stated that variability is measured by means of fluctuations and may be analyzed in terms of dynamics. A separation into deterministic and stochastic dynamics may correspond to a discrimination between external forces, acting as noise on the studied system, and the intrinsic response to them. This is in line with the characterization of a stochastic process in terms of a slowly and a fast varying part. For the first example of appli-

cation, i.e. the power performance modelling for a wind turbine in Chapter 3, these external forces are predominantly the wind fluctuations the turbine is exposed to. Within the scope of the stochastic modelling, I have primarily not been interested in a characterization of these fluctuations but rather how they are transferred to the power output of the wind turbine. Accordingly, I have concentrated on the reconstruction of the drift dynamics reflecting the internal behaviour of the studied system. This has been different for the modelling of the balance performance in Chapter 4. The used balance board together with the test subject on top of it forms a complex system that cannot be described in all details but is instead modelled on the whole, and the stochastic short-term fluctuations arise within the system itself. To evaluate the total balance performance, I have considered the estimated values for the drift as well as for the diffusion coefficient. The Langevin approach may accordingly give for very different systems a plausible framework. However, the significance of the obtained drift and diffusion coefficients, i.e. the reliability of the respective reconstructed values, has to be evaluated for each application separately. In this thesis, I have proposed several procedures for this kind of evaluation that may serve as very useful tools for future applications as well as for the revision of already investigated systems.

Danksagung

Mein Dank gilt all denjenigen, die mich während meiner Promotion unterstützt haben und diese Arbeit so erst möglich gemacht haben. Bedanken möchte ich mich im Speziellen bei Prof. Dr. Joachim Peinke für die Betreuung, aber insbesondere auch für den Freiraum beim Erstellen dieser Arbeit. Ein großer Dank gilt auch Prof. Dr. Volker Lippens für die gute Kooperation, die vielen Anregungen und Nachfragen. Bei Prof. Dr. Ulrike Feudel möchte ich mich für die Zeit bedanken, die sie sich für das Begutachten der vorliegenden Arbeit genommen hat.

Weiterhin danke ich meinen Kollegen, Freunden und Bekannten, die mich während der letzten Jahre begleitet haben, für eine gute und Zeit. Namentlich erwähnen möchte ich insbesondere Sarah Biela, Bernhard Stoevesandt und Matthias Wächter, die sich mit den vorläufigen Versionen dieser Arbeit abgemüht und mir mit ihren Korrekturen, Anmerkungen und Fragen sehr geholfen haben.

Besonderer Dank gilt zuletzt meiner Familie, im engen und im weiten Sinn, und meinem Fabian, der die letzten Wochen und Monate mit mir zusammen durchgestanden hat – zuletzt auch der kleinen schwarzen Katze, die sich jetzt schon wieder auf meine Maus gesetzt hat.

Lebenslauf

Persönliche Daten:

Name	Julia Gottschall
Geburtsdatum	14.03.1982
Geburtsort	Göttingen
Nationalität	Deutsch

Schulausbildung:

1988 – 1992	Grundschule Hardeggen
1992 – 1994	Orientierungsstufe Moringen
1994 – 2001	Gymnasium Corvinianum, Northeim (Abitur)

Studium:

10.2001 – 04.2003	Grundstudium im Fach Physik (Vordiplom), Universität Heidelberg
05.2003 – 09.2005	Hauptstudium im Fach Physik, Universität Heidelberg
08.2004 – 07.2005	Auslandsstudium / Erasmus-Stipendium, Universität Uppsala (Schweden) Masterarbeit (M. Sc.) am Department für Theoretische Physik

Wissenschaftliche Arbeit:

seit 09.2005	Promotion am Institut für Physik, Universität Oldenburg; wissenschaftliche Mitarbeiterin bei ForWind – Zentrum für Windenergieforschung
--------------	--

Erklärung

Ich erkläre hiermit, dass

- (a) eine Promotion zum Dr. rer. nat. angestrebt wird,
- (b) ich diese Arbeit selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe, und dass die Arbeit in Teilen bereits veröffentlicht bzw. zur Veröffentlichung eingereicht wurde,
- (c) diese Dissertation weder in ihrer Gesamtheit noch in Teilen einer anderen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegen hat.

Oldenburg, den 28. Oktober 2008

.....

(Julia Gottschall)

Publikationsliste

Teile der vorliegenden Arbeit bestehen aus den folgenden bereits veröffentlichten oder zum Zeitpunkt der Fertigstellung eingereichten wissenschaftlichen Artikeln:

J. GOTTSCHALL UND J. PEINKE: On the definition and handling of different drift and diffusion estimates. In: *New Journal of Physics* 10, 083034 (2008).

J. Peinke hatte bei dieser Arbeit eine betreuende Funktion. Die wissenschaftliche Ausarbeitung stammt von J. Gottschall.

J. GOTTSCHALL AND J. PEINKE: How to improve the estimation of power curves for wind turbines. In: *Environmental Research Letters* 3, 015005 (2008).

J. Peinke hatte bei dieser Arbeit eine betreuende Funktion. Die wissenschaftliche Ausarbeitung stammt von J. Gottschall.

J. GOTTSCHALL, J. PEINKE, V. LIPPENS, AND V. NAGEL: Exploring the dynamics of balance data – movement variability in terms of drift and diffusion. Eingereicht bei: *Physics Letters A* (2008).

J. Peinke hatte bei dieser Arbeit eine betreuende Funktion. V. Lippens and V. Nagel stellten die gemessenen Daten zur Verfügung und gaben Anregungen zur verwendeten Analyseverfahren. Die wissenschaftliche Ausarbeitung stammt von J. Gottschall.

Die Anteile der aufgeführten Autoren entsprechen den Angaben.

Oldenburg, den 28. Oktober 2008

.....

(Prof. Dr. J. Peinke)