Quantitative Estimation of Drift and Diffusion Functions from Time Series Data

David Kleinhans and Rudolf Friedrich

Westfälische Wilhelms-Universität Münster, Institut für Theoretische Physik, 48149 Münster, Germany

Summary. This contribution provides an introduction to the concept of drift and diffusion functions for complex dynamical systems such as wind energy converters. These functions easily can be estimated from measured data. However, one has to be aware about intrinsic errors in the estimation procedure that are discussed in the following.

1 Introduction

Researchers in the field of the construction of wind energy converters are confronted with a complex problem: the number of degrees of freedom of the wind turbine is extraordinary high. In addition to the adjustable parameters of the rigid body such as the pitch of the rotor-blades many dynamical modes of different parts of the converters have to be considered for a complete description. Moreover the incoming flow is turbulent and fluctuating in space as well as in time.

Complex behaviouriour in systems far from equilibrium can quite often be traced back to rather simple laws due to the existence of processes of self-organization. For adequate order parameters the dynamics is determined by stochastic differential equations incorporating deterministic as well as stochastic forces. Knowledge of the deterministic part of the dynamics can lead to a deeper understanding of the properties of the system under consideration while the stochastic forces account for the effects of the fluctuating microscopic degrees of freedom. For certain order parameters these forces have properties that are well-known in the theory of stochastic processes. Imagine for example the power output of a wind turbine. Usually the power output is investigated as a function of the (mean) wind speed. Without a doubt much more parameters of the turbine effect the output power and the dynamics of the system act much faster than the common averaging periods.

Recently it has become evident, that knowledge of the stochastic dynamics has significant advantages with respect to the conventional wind power curves:

Anahua *et al.* considered the power output of the turbine as a stochastic process [2]. A standard method allows for direct estimation of the characteristic drift and diffusion functions from measured data [1]. By means of this method Anahua *et al.* could extract the real time dynamics and in the meantime are able to detect small deviations of the control system of the turbine from the optimal working state.

2 Direct Estimation of Drift and Diffusion

Generally one has to distinguish between dynamical and measurement noise: measurement noise is superimposed to the data during the measurement process and has no further influence on the system's dynamics. On the other hand many complex systems on a macroscopic scale show some intrinsic, dynamical noise stemming from the microscopic degrees of freedom. Under certain conditions the time evolution of the state $\boldsymbol{x}$ of such systems can be described by Langevin equations of the type

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

$\boldsymbol{\Gamma}$ represents an independent, delta correlated and normal distributed stochastic force that obeys $\langle \Gamma_i(t)\Gamma_j(t')\rangle = 2\delta_{ij}\delta(t-t')$. We apply Itô's interpretation of stochastic integrals [3]. The corresponding Fokker-Planck equation (FPe) characterizes the evolution of the probability density function (pdf) f with time,

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

$\boldsymbol{D}^{(1)}(\boldsymbol{x})$ is called the drift vector, $D^{(2)}(\boldsymbol{x})$ the diffusion matrix of the corresponding stochastic system.

From the Kramers-Moyal expansion [3] – the more general origin of the Fokker-Planck equation that covers non-markovian processes – the following definition is known:

$$D^{(n)}(\boldsymbol{x}) := \frac{1}{n!} \lim_{\tau \to 0} \frac{1}{\tau} \left\langle [\boldsymbol{x}(t+\tau) - \boldsymbol{x}(t)]^n \,|\boldsymbol{x}(t) = \boldsymbol{x} \right\rangle \quad . \tag{3}$$

It has been shown [1] that this expression applied for $n = 1$ and $n = 2$ can be used for direct estimation of drift and diffusion functions respectively from time series data. This procedure successfully has been applied to the power output of wind turbines [2] and various problems in medical and life science. The computational requirements for this method are outstandingly low. However, the required discretization of state space and – in particular – the limiting procedure with respect to the time increment makes high demands on the time series data with regard to the sampling frequency and the amount of data-points. From now on the estimation of an one-dimensional process is discussed. The generalization to higher dimensions follows accordingly.

3 Stability of the limiting procedure

Any measured time series has a finite sampling rate that limits the available time increments for the limiting procedure (3). Hence this expression has to be extrapolated to the value $\tau \equiv 0$.

A formal solution of the FPe for the conditional pdf $p(x, t|x_0, t_0)$ is

$$p(x, t|x_0, t_0) = \exp\left[\hat{L}\,(t - t_0)\right] \delta(x - x_0) \tag{4}$$

with $\hat{L}$ being the Fokker-Planck operator. A Taylor expansion of this expression yields

$$p(x, t_0 + \tau|x_0, t_0) = \left(1 + \tau\hat{L} + \frac{\tau^2}{2}\hat{L}^2 + \mathcal{O}(\tau^3)\right) \delta(x - x_0) \quad . \tag{5}$$

This pdf can be used for analytical calculation of the conditional moments (3). Eventually one can assess the deviations of the estimate of drift and diffusion $D_E^{(i)}(x, \tau)$ for finite τ from the intrinsic functions $D^{(i)}(x)$. The first order corrections read:

$$D_E^{(1)}(x, \tau) \approx D^{(1)}(x) + \frac{\tau}{2}\left[D^{(1)}(x)\frac{\partial}{\partial x}D^{(1)}(x) + D^{(2)}(x)\frac{\partial^2}{\partial x^2}D^{(1)}(x)\right]$$

$$D_E^{(2)}(x, \tau) \approx D^{(2)}(x) + \frac{\tau}{2}\left[D^{(1)}(x)D^{(1)}(x) + 2D^{(2)}(x)\frac{\partial}{\partial x}D^{(1)}(x)\right. \tag{6a}$$

$$\left. + D^{(1)}(x)\frac{\partial}{\partial x}D^{(2)}(x) + D^{(2)}(x)\frac{\partial^2}{\partial x^2}D^{(2)}(x)\right] \quad . \tag{6b}$$

Depending on the shape of drift and diffusion functions significant deviations from the intrinsic functions occur for finite τ. These deviations cannot be grasped with statistical considerations as they originate in the properties of the propagator for finite time.

4 Finite length of time series

On the other hand one has to consider the finite number of data points. Discretization of state space confines the number of points even more. Especially in sparsely populated regions that come along with natural boundary conditions the low density leads to huge errors in the estimation of drift and diffusion.

A suitable measure for the error margin of the estimate is the standard deviation, the root mean square displacement of the increments from their mean. If N measurements contribute to the averages, the resulting error $E_N\left[D_E^{(1)}(x_0, \tau)\right]$ in the estimated drift function gets

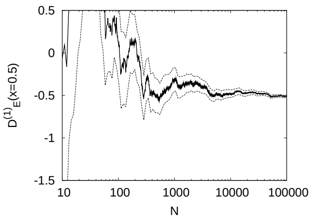

Fig. 1. Estimated drift function $D_E^{(1)}(x = 0.5)$ of synthetic Ornstein-Uhlenbeck process as function of N. Dashed region marks the symmetric error-bars corresponding to (7).

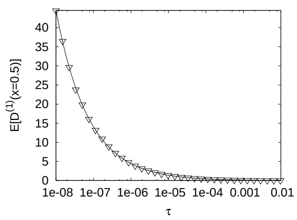

Fig. 2. Error of drift estimate as function of time increment τ (triangles). A divergent behaviour for $\tau \to 0$ is evident. The solid line represents the best fit $f(\tau) = 0.0045/\sqrt{\tau}$.

$$E_N\left[D_E^{(1)}(x_0,\tau)\right] = \sqrt{\frac{2}{\tau}\frac{D_E^{(2)}(x_0,\tau)}{N} - \frac{\left[D_E^{(1)}(x_0,\tau)\right]^2}{N}} \quad . \tag{7}$$

In sum the statistical error in the estimated drift function is proportional to $(N\tau)^{-1/2}$. Figures 1 and 2 illustrate the divergent behaviour of the estimated drift coefficient $D_E^{(1)}$ in the cases of few data-points and small time increments considering as example data from an Ornstein-Uhlenbeck process.

5 Conclusion

In conclusion there is simple method to estimate the dynamical drift and diffusion functions from measured data. This method can be used to describe the dynamical behaviour of complex systems such as wind energy converters. Quantitative results from this method have to be considered carefully for the reasons discussed in sections 3 and 4.

We would like to stress that there is a more recent extension that avoids the evaluation of the conditional moments in the limit of small time increments and improves the accuracy of the results substantially [4].

References

1. S. Siegert, R. Friedrich, and J. Peinke, Physics Letters A **243**, 275 (1998).
2. E. Anahua, F. Böttcher, S. Barth, and J. Peinke, Proceedings of the European Wind Energy Conference (EWEC), London, UK (2004).
3. H. Risken, *The Fokker-Planck equation*, 2nd ed. (Springer-Verlag, Berlin, 1989).
4. D. Kleinhans, R. Friedrich, A. Nawroth, and J. Peinke, Phys Lett A **346**, 42 (2005).