In: **Chaos in Brain?**, Proc. of the 1999 Workshop, K. Lehnertz, J. Arnhold, P. Grassberger and C. E. Elger, eds., (World Scientific, Singapore, 2000), p. 226 - 229.

ANALYSIS OF FLUCTUATING DATA SETS OF DIFFUSION PROCESSES

SILKE SIEGERT AND RUDOLF FRIEDRICH

Institute for Theoretical Physics 3, University of Stuttgart, Pfaffenwaldring 57, D-70550 Stuttgart, Germany

E-mail: silke@theo3.physik.uni-stuttgart.de, fiddi@theo3.physik.uni-stuttgart.de

PETER HÄNGGI

Institute for Theoretical Physics 1, University of Augsburg, Universitätsstr. 1, D-86135 Augsburg, Germany

E-mail: Hanggi@Physik.Uni-Augsburg.de

An algorithm for analyzing stochastic data sets of the diffusion type in order to find drift and diffusion coefficients of the affiliated Fokker-Planck equation is presented. The results of an application of the method to data sets of one- and two-dimensional, stationary and nonstationary systems are discussed.

1 Introduction

Measurements concerning biological, economical, physical or technical systems deliver noisy data sets. Looking at these data sets it is of great interest to find the deterministic and stochastic parts of the dynamics.

In this contribution an algorithm is presented that allows the analysis of fluctuating time series data originating from diffusion processes. Appropriate rules for the dynamics of the investigated system can be extracted solely from the given data sets, therefore the method may be called a *data-driven* approach. However, it has to be emphasized that not all stochastic processes in general, but at least diffusion processes (for definition see e.g. [1]) can be investigated by an application of the algorithm.

2 Diffusion processes

The dynamics of diffusion processes can be formulated either by a Langevin equation (1) or by a Fokker-Planck equation (2) (see e.g. [2]).

The Langevin equation evaluates the time derivative of the d-dimensional stochastic variable $\mathbf{X}(t)$ as sum of a deterministic part $\mathbf{h}(\mathbf{X}(t))$ and a noise term $\mathbf{g}(\mathbf{X}(t))\mathbf{\Gamma}(t)$, where $\mathbf{\Gamma}(t)$ is a white noise factor with $<\Gamma_i(t)\Gamma_j(s)>=\delta_{ij}\delta(t-s)$:

$$\frac{d}{dt}X_i(t) = h_i(\mathbf{X}(t)) + \sum_{j=1}^{d} g_{ij}\Gamma_j(t). \tag{1}$$

In the following Stratonovich's interpretation of the Langevin equation and its stochastic integrals is adopted (see [3]).

Alternatively, the dynamic behaviour of the stochastic process can be described by a partial differential equation for the conditional probability density $p(\mathbf{x}, t \,|\, \mathbf{y}, \tau)$, where $\mathbf{x}, \mathbf{y}$ belong to the state space of the stochastic variable $\mathbf{X}(t)$:

$$\frac{\partial p(\mathbf{x},t\,|\,\mathbf{y},\tau)}{\partial t} = \left(-\sum_{i=1}^{d} \frac{\partial}{\partial x_i} D_i^{(1)}(\mathbf{x},t) + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\mathbf{x},t) \right) p(\mathbf{x},t\,|\,\mathbf{y},\tau). \tag{2}$$

$\mathbf{D}^{(1)}(\mathbf{x},t)$ and $\mathbf{D}^{(2)}(\mathbf{x},t)$ are called drift and diffusion coefficients of the Fokker-Planck equation. If these coefficients are found by data analysis for an investigated diffusion process, the dynamic behaviour of the system is completely known. The coefficients of the Fokker-Planck equation, $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, are connected with the deterministic and stochastic parts $\mathbf{h}$ and $\mathbf{g}$ of the Langevin equation by the following relations

$$D_i^{(1)}(\mathbf{x},t) = h_i(\mathbf{x},t) + \frac{1}{2} \sum_{l,m=1}^{d} g_{lm}(\mathbf{x},t) \frac{\partial}{\partial x_l} g_{im}(\mathbf{x},t) \tag{3}$$

$$D_{ij}^{(2)}(\mathbf{x},t) = \sum_{l=1}^{d} g_{il}(\mathbf{x},t) g_{jl}(\mathbf{x},t). \tag{4}$$

3 Algorithm for analyzing data of diffusion processes

The primary task within the analysis of fluctuating data originating from diffusion processes is to find drift and diffusion coefficients ([4], [5]). They are defined by the following relations

$$D_i^{(1)}(\mathbf{x},t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \langle X_i(t+\Delta t) - x_i \rangle \Big|_{\mathbf{X}(t)=\mathbf{x}}, \tag{5}$$

$$D_{ij}^{(2)}(\mathbf{x},t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \langle (X_i(t+\Delta t) - x_i)(X_j(t+\Delta t) - x_j) \rangle \Big|_{\mathbf{X}(t)=\mathbf{x}}. \tag{6}$$

First, the state space of the investigated system, represented by the data of the regarded time series, has to be discreticized. For stationary processes drift and diffusion coefficients are no longer time-dependent. Therefore, the mean values in expression (5), (6) can be determined numerically for small but finite values of Δt. Because the limit $\Delta t \to 0$ can never be reached exactly, a higher order relation can be used for the diffusion term:

$$\begin{aligned} D_{ij}^{(2)}(\mathbf{x},t) = {} & \frac{1}{\Delta t} \Big\langle (X_i(t+\Delta t) - x_i)(X_j(t+\Delta t) - x_j) \Big\rangle \Big|_{\mathbf{X}(t)=\mathbf{x}} \\ & - \frac{1}{\Delta t} \langle X_i(t+\Delta t) - x_i \rangle \Big|_{\mathbf{X}(t)=\mathbf{x}} \cdot \langle X_j(t+\Delta t) - x_j \rangle \Big|_{\mathbf{X}(t)=\mathbf{x}} \end{aligned} \tag{7}$$

For nonstationary processes, an analysis window has to been drawn along the whole time series. Within one window the process is assumed to be stationary. Therefore, for this time interval drift and diffusion coefficients can be estimated according to the way described above. By regarding the analysis results of temporal overlapping windows the nonstationarity of the dynamics can be reconstructed. The time dependence of the dynamic behaviour of deterministic and stochastic parts can be classified in three groups: stepwise change, continuous change, and fast periodic change.

If requested, analytic functions may be adapted to the numerically found values of drift and diffusion coefficients.

4 Application of the algorithm to artificial data sets

4.1 Van-der-Pol oscillator

In fig. 1 extracts from the time series of a Van-der-Pol oscillator with dynamical noise can be seen. The time series originate from an integration of the Langevin equations

$$\frac{d}{dt}X_1(t) = X_2(t), \qquad \frac{d}{dt}X_2(t) = X_2(0.3 - X_1^2) - X_1 + 0.3\Gamma(t), \tag{8}$$

where $\Gamma(t)$ is a Gaussian white noise term. These data sets have been analyzed by the presented algorithm. In fig. 2 a vector plot of the two-dimensional drift coefficient can be seen. Additionally, a trajectory that has been integrated according to the numerically determined drift coefficient field and the affiliated trajectory with the same starting point, found according to the deterministic part of the differential equations (8), have been plotted. A quite good conformity can be recognized.

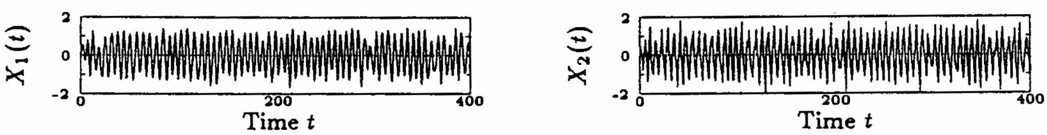

Figure 1. Artificial time series of system (8): Stochastic variables $X_1(t)$, $X_2(t)$ over time t in arbitrary units.

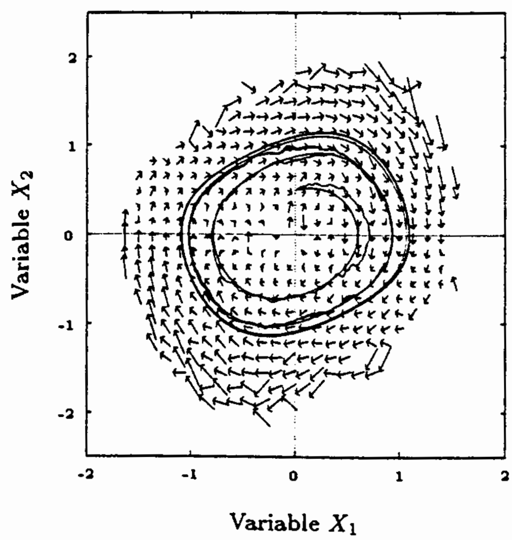

Figure 2. Vector field of the 2-dimensional drift coefficient $D^{(1)}$, numerically determined for the time series in fig. 1. The axis refer to the state space of the stochastic variables X_1, X_2.

4.2 *Pitchfork bifurcation*

As an example for the analysis of a nonstationary diffusion process a pitchfork bifurcation has been investigated. The time series in fig. 3 consists of three different parts according to the Langevin equations

$$
\begin{aligned}
t \in (0, 8000): &\quad \dot{X}(t) = 0.1X(t) - X^3(t) + 0.05\Gamma(t), \\
t \in (8000, 16000): &\quad \dot{X}(t) = -X^3(t) + 0.05\Gamma(t), \\
t \in (16000, 24000): &\quad \dot{X}(t) = -0.1X(t) - X^3(t) + 0.05\Gamma(t).
\end{aligned} \tag{9}
$$

An analysis window with a size of 4000 time units has been moved along the whole time series. If the window regards only data from one of the three segments, the analysis algorithm delivers the expected results (fig. 4). If an analysis window covers a part of the time series, that contains two different states of the nonstationary dynamics, the numerical procedure for determining drift and diffusion terms will deliver weighted average values of the exact coefficients.

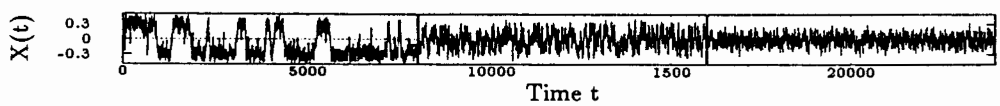

Figure 3. Nonstationary time series of system (9): Variable $X(t)$ over time t in arbitrary units.

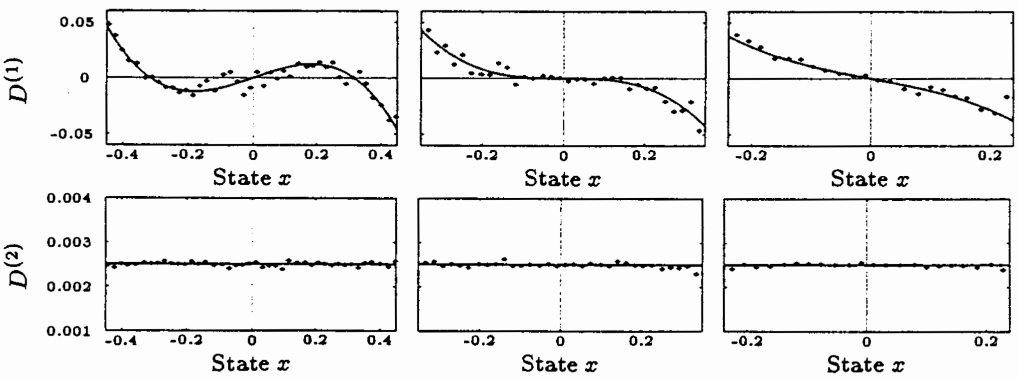

Figure 4. Numerically determined drift and diffusion coefficients $D^{(1)}$ and $D^{(2)}$ (points) over state x for the time series in fig. 3. The solid lines depict the expected functions of the coefficient. The 3 subfigures refer to the 3 different segments of the time series.

References

1. L. Arnold, *Stochastische Differentialgleichungen*, R. Oldenbourg Verlag München, Wien, 1973
2. P. Hänggi and H. Thomas, *Phys. Rep.* **88**, 207 (1982).
3. R. L. Stratonovich, *SIAM J. Control* 4, 362 (1966)
4. S. Siegert, R. Friedrich and J. Peinke, *Phys. Lett. A* **243**, 275 (1998).
5. R. Friedrich and J. Peinke, *Phys. Rev. Lett.* **78**, 863 (1997).