



Extracting model equations from experimental data

R. Friedrich ^{a,*}, S. Siegert ^a, J. Peinke ^b, St. Lück ^b, M. Siefert ^b, M. Lindemann ^{c,d},
J. Raethjen ^d, G. Deuschl ^d, G. Pfister ^c

^a *Institute for Theoretical Physics, University of Stuttgart, D-70550 Stuttgart, Germany*

^b *Department of Physics, University of Oldenburg, D-26111 Oldenburg, Germany*

^c *Institute for Experimental and Applied Physics, University of Kiel, D-24098 Kiel, Germany*

^d *Department of Neurology, University of Kiel, D-24105 Kiel, Germany*

Received 18 April 2000; accepted 2 May 2000

Communicated by J. Flouquet

Abstract

This letter wants to present a general data-driven method for formulating suitable model equations for nonlinear complex systems. The method is validated in a quantitative way by its application to experimentally found data of a chaotic electric circuit. Furthermore, the results of an analysis of tremor data from patients suffering from Parkinson's disease, from essential tremor, and from normal subjects with physiological tremor are presented, discussed and compared. They allow a distinction between the different forms of tremor. © 2000 Elsevier Science B.V. All rights reserved.

PACS: 05.10.Gg; 05.45.-a; 05.40.Ca

1. Introduction

A basic aim of scientific research is to set up reasonable models for considered systems. A suitable model should reproduce the observed quantities and help to gain a deeper understanding of the system. Usually, collected data and known properties of the system, as symmetry relations for example, serve as a basis for the modelling. In contrast to this a *general data-driven way* for formulating suitable model equations for nonlinear complex systems is presented in the following.

An important and wide class of dynamic systems can be described by the following differential equation

$$\frac{d}{dt}X(t) = \underbrace{g(X(t),t)}_{\text{deterministic part}} + \underbrace{h(X(t),t)\Gamma(t)}_{\text{stochastic part}}, \quad (1)$$

the Langevin equation [1,2]. $X(t)$ denotes the time dependent d -dimensional stochastic vector which characterises the system completely. The time derivative of $X(t)$ can be expressed as sum of a deterministic part g and a stochastic part $h \cdot \Gamma$,

* Corresponding author. Fax: +49-711-685-5098.
E-mail address: fiddi@theo3.physik.uni-stuttgart.de
(R. Friedrich).

where $\Gamma(t)$ stands for terms of δ -correlated Gaussian white noise and where the $d \times d$ -matrix \mathbf{h} fixes the dynamic influence of the noise on the system. For the functionals \mathbf{g} and \mathbf{h} no further assumptions have to be made; \mathbf{g} can be nonlinear, and therefore also deterministic chaos can be formulated by a Langevin Eq. (1). The investigation of complex systems has shown the necessity to find a description of a system by Langevin equations directly from measured data sets [3].

For this wide class of dynamic systems (1) a general method for finding the deterministic and stochastic laws solely by data analysis will be presented in this letter. If this only condition, the describability of the system's dynamics by an evolution equation like (1), is given, no further assumptions or pre-knowledge have to be included in the following analysis. Deterministic and noisy parts of the dynamics can be separated and quantified, and model equations for the dynamics can be set up by the data-driven method.

2. Numerical method

The considered class of dynamic systems (1) is characterised by the missing of memory effects, i.e. for the time development of the system we need to know only the state of one vector $\mathbf{X}(t)$ at a given time t and not its evolution in the past. A system with this quality is called a Markovian system [4]. The conditional probability density $p(\mathbf{x}_{n+1}, t + \tau | \mathbf{x}_n, t; \mathbf{x}_{n-1}, t - \tau; \dots)$ describes the probability of states \mathbf{x}_{n+1} of the system's variable \mathbf{X} at time $t + \tau$ under the condition that the system is in state \mathbf{x}_n at time t , has been in state \mathbf{x}_{n-1} at time $t - \tau$ and so on. The Markovian property of a system can be tested by comparing the multiple conditional probability densities with the one-step conditional probability densities $p(\mathbf{x}_{n+1}, t + \tau | \mathbf{x}_n, t)$. If both expressions agree the time development of the probability density depends only on the present state and not on its evolution in the past.

The assumed qualities of the driving noise terms Γ as being Gaussian white noise functions can be validated, as well, by looking at the conditional probability density distributions.

The central ideas of the presented algorithm will be explained in the following: First, stationary dynamics shall be assumed, i.e. the deterministic and stochastic parts \mathbf{g} and \mathbf{h} are not explicitly time dependent¹. Every time t_i , the system's trajectory meets an arbitrary but fixed point \mathbf{x} in state space, the localisation of the trajectory at time $t_i + \tau$ is determined by the deterministic function $\mathbf{g}(\mathbf{x})$, which is constant for fixed \mathbf{x} , and by the stochastic function $\mathbf{h}(\mathbf{x})\Gamma(t_i)$ with constant \mathbf{h} for fixed \mathbf{x} and Gaussian distributed white noise $\Gamma(t)$. With this background in mind, the following relationships which have been proved in a strict mathematical way [5,6] using Itô's definitions for stochastic integrals [7], become understandable:

$$\mathbf{g}(\mathbf{x}) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle \mathbf{X}(t + \tau) - \mathbf{x} \rangle |_{\mathbf{X}(t) = \mathbf{x}} \quad (2)$$

$$\mathbf{h}(\mathbf{x}) \cdot \mathbf{h}^T(\mathbf{x}) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle (\mathbf{X}(t + \tau) - \mathbf{x}) \times (\mathbf{X}(t + \tau) - \mathbf{x})^T \rangle |_{\mathbf{X}(t) = \mathbf{x}} \quad (3)$$

Under the condition that the system's trajectory meets the point \mathbf{x} at time t , i.e. $\mathbf{X}(t) = \mathbf{x}$, the deterministic part \mathbf{g} of the dynamics can be evaluated for small τ by the difference of the system's state at time $t + \tau$ and the state at time t , averaged over an ensemble, or in the regarded stationary case, averaged over all $t = t_i$ of the whole time series with $\mathbf{X}(t_i) = \mathbf{x}$. The limit $\tau \rightarrow 0$ can be reached by extrapolation. In a similar way, the stochastic influences can be determined, regarding quadratic terms in the averages [8]. For every point \mathbf{x} in state space, that is visited statistically often by the trajectory, deterministic and stochastic parts of the dynamics can be estimated numerically. A schematic illustration of this procedure is shown in Fig. 1.

As final step, analytic functions can be fitted to the numerically determined values of $\mathbf{g}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$

¹ Non-stationary dynamic systems, i.e. with functionals \mathbf{g} and \mathbf{h} being explicitly time dependent, can be investigated by a moving window technique. An analysis window is drawn along the whole time series in overlapping steps. In each window, the system's dynamics is assumed to be stationary. When interpreting the results, the quality of the algorithm of calculating weighted averages of the real functionals has to be taken in mind.

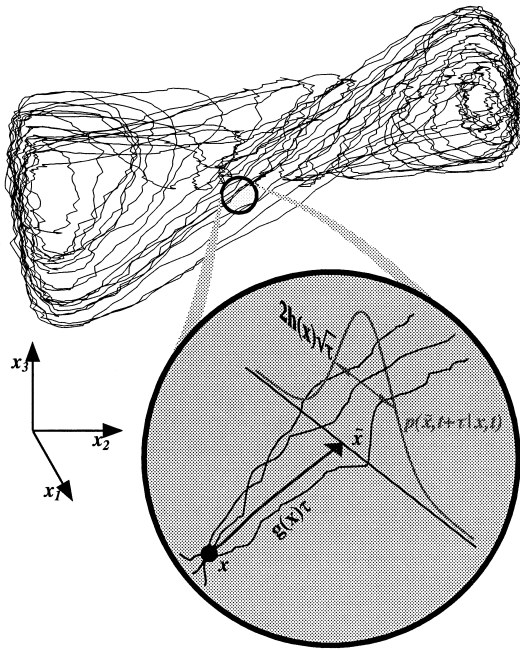


Fig. 1. Schematic illustration of the presented algorithm for analysing noisy data sets and calculating the deterministic and stochastic parts of the underlying dynamics. In the upper part of the figure a trajectory of a chaotic oscillator (Shinriki) with dynamical noise is shown. From the trajectory all parts are selected which run exactly or within some limits at any time t_i through a certain point x in state space, i.e. $X(t_i) = x + \Delta x$. The distribution of the values \bar{x} taken at the next step $X(t_i + \tau)$ of these trajectory parts is given by a Gaussian function with mean $x + g(x)\tau$ and standard deviation $h(x)\sqrt{\tau}$, where g and h are defined by Eq. (1). $g(x)$ is the mean velocity of the trajectory at point x , $h(x)$ represents the fluctuations of the velocity at this point. The conditional probability density $p(\bar{x}, t + \tau | x, t)$ can be evaluated by this way.

in order to formulate model equations for the investigated system.

3. Analysis of electronic data

Next, the application of the method to data sets from two experimental systems will be presented. As first example, a chaotic electric circuit has been chosen. Its dynamics is formed by a damped oscillator with nonlinear energy support and additional dynamic noise terms. In this case, well defined electric quantities are measured for which the dynamic equations are known. The measured time series are analysed according to the numerical algorithm de-

scribed above. Afterwards, the numerically determined results and the expected results according to the system's evolution equations are compared.

The electric circuit is shown as schematic illustration in Fig. 2. Its dynamic equations are given by the following evolution equations, where the deterministic part is known as Shinriki oscillator [9]:

$$\begin{aligned} \dot{X}_1 &= \left(-\frac{1}{R_{NIC}C_1} - \frac{1}{R_1C_1} \right) X_1 - \frac{f(X_1 - X_2)}{C_1} \\ &\quad + \frac{1}{R_{NIC}C_1} \Gamma(t) \\ &= g_1(X_1, X_2) + h_1 \Gamma(t) \end{aligned} \tag{4}$$

$$\begin{aligned} \dot{X}_2 &= \frac{f(X_1 - X_2)}{C_2} - \frac{1}{R_3C_2} X_3 \\ &= g_2(X_1, X_2, X_3) \end{aligned} \tag{5}$$

$$\begin{aligned} \dot{X}_3 &= -\frac{R_3}{L} (X_2 - X_3) \\ &= g_3(X_2, X_3). \end{aligned} \tag{6}$$

X_1, X_2 and X_3 denote voltage terms, R_i are values of resistors, L_i and C_i stand for inductivity and capacity values. The function $f(X_1 - X_2)$ denotes the characteristic of the nonlinear element. The quantities X_i , characterising the stochastic variable of the Shinriki oscillator with dynamical noise, were measured by means of a 12 bit A/D converter. Our analysis is based on the measurement of 100.000 data points, the affiliated trajectory is shown in the upper part of Fig. 1.

The measured three-dimensional time series were analysed according to the algorithm described above. The determined deterministic dynamics – expressed

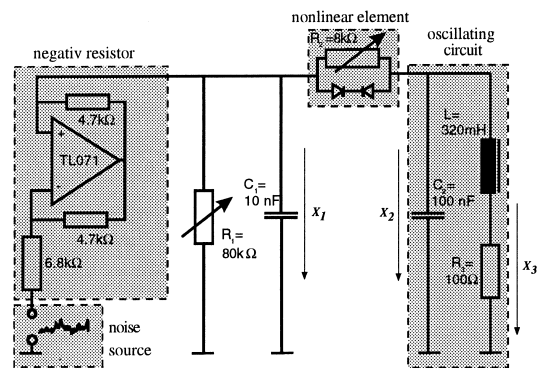


Fig. 2. Illustration of the investigated electric circuit. The structure of the dynamics is a chaotic oscillator with dynamical noise.

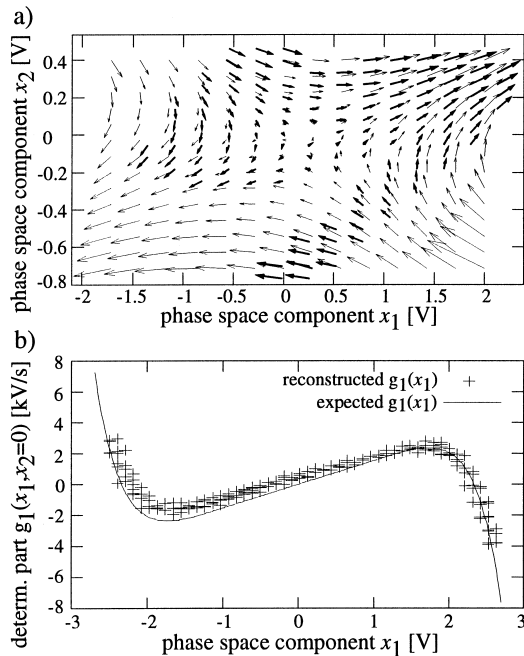


Fig. 3. Cuts of the function $g(x)$ reconstructed from experimental data of the electric circuit (illustrated in Fig. 2) in comparison with the expected functions according to the known differential Eqs. (4)–(6). In part (a) the cut $(g_1(x_1, x_2), g_2(x_1, x_2, x_3 = 0))$ is shown as two-dimensional vector field. Thick arrows represent values determined by data analysis, thin arrows represent the theoretically expected values. In areas of the state space where the trajectory did not go during the measurement no estimated values for the functions exist. Figure (b) shows the one dimensional cut $g_1(x_1, x_2 = 0)$. Points represent values estimated numerically by data analysis. Additionally, the affiliated theoretically curve is printed, as well.

by the deterministic part of the evolution equations – corresponds to a vector field in the three-dimensional state space. For presentation of the results, cuts of lower dimension have been generated. Part (a) of Fig. 3 illustrates the vector field

$$(g_1(x_1, x_2), g_2(x_1, x_2, x_3 = 0))$$

of the reconstructed deterministic parts affiliated with (4), (5). Furthermore, the one-dimensional curve

$g_1(x_1, x_2 = 0)$ is drawn in part (b). Additionally to the numerically determined results, found by data analysis, the expected vector field and curve (Eqs. (4), (5)) are shown for comparison. A good agreement can be recognized.

4. Analysis of tremor time series

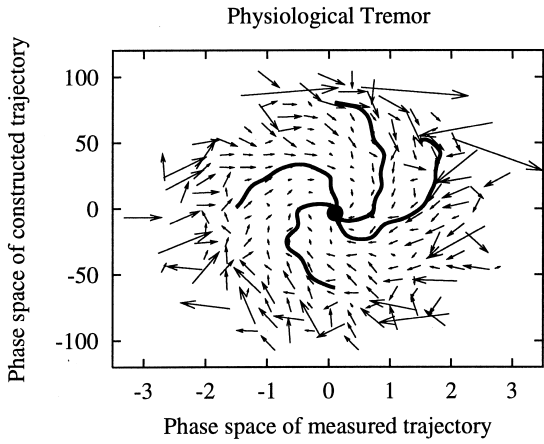
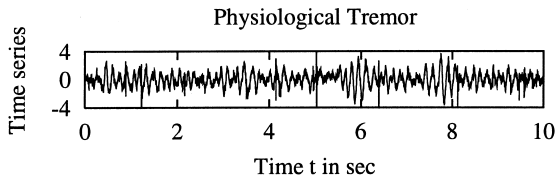
In the second example, data originating from neurophysiology are investigated. Time series of hand tremor have been recorded in normal subjects with physiological tremor (PT), in patients with essential tremor (ET) and in patients suffering from Parkinson's disease (PD). The data sets were measured by a lightweight piezoresistive accelerometer with a sampling rate of 800 Hz. The outstretched hand was supported at the wrist. For the analysis 24 000 data points per experiment were used. The data sets were scaled to variance 1.

It is assumed that the underlying dynamics of the measured time series $X_1(t)$ can be expressed by a two-dimensional Langevin equation (1). Using a time delay method a second time series $X_2(t)$ was created out of the measured time series $X_1(t)$. Afterwards, the algorithm described above was applied. Up to now 20 data sets have been investigated, measured at patients whose disease can be clearly classified within one of the three groups: physiological tremor, essential tremor and Parkinson's disease. Further data sets will be investigated, but already now a clear structure of the dynamic behaviour appears. Typical results of the analysis for the three tremor groups together with extracts of the investigated time series can be seen in Fig. 4.

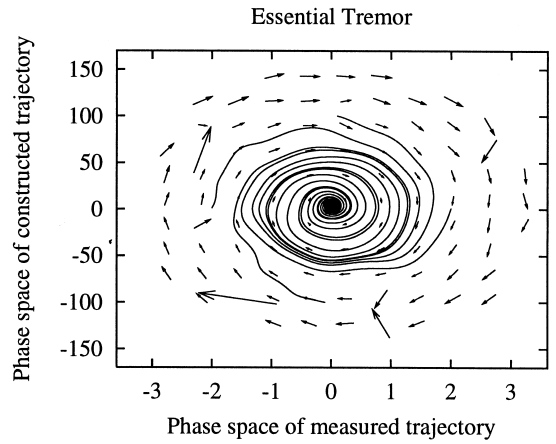
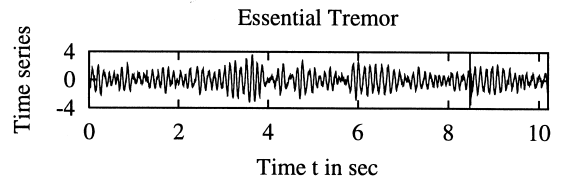
Whereas the deterministic parts of the dynamics in the case of physiological tremor can be described as a fixed point with large damping and weak rotation, the deterministic parts of the dynamics in the case of essential tremor are characterised by a fixed

Fig. 4. Typical results of an analysis of tremor data achieved by the presented algorithm. In part (a) variance normalised accelerometer data from a normal subject with physiological tremor was investigated, in part (b) from a patient with essential tremor and in part (c) from a patient suffering from Parkinson's disease. In every part a subsequence of the used time series and the numerical results for the deterministic parts of the dynamics, drawn as vector diagram, are shown. The curves are trajectories integrated along these vector fields. Whereas in part (a) and (b) the underlying deterministic parts of the total dynamics can be described as fixed points the deterministic parts of the dynamics in the case of Parkinson's disease are characterised as periodic limit cycle. The dynamics of physiological and essential tremor can be separated by their damping, i.e. the number of cyclings of the trajectory on its way to the fixed point.

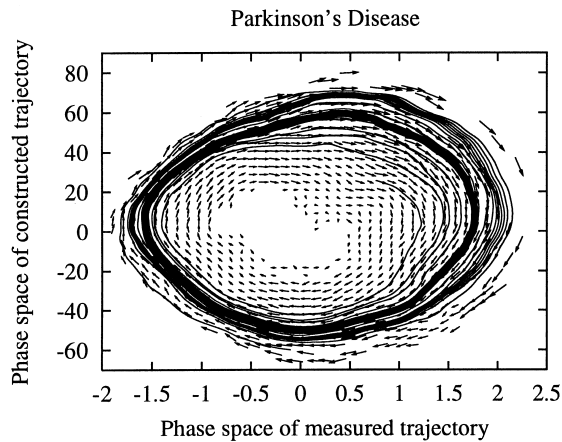
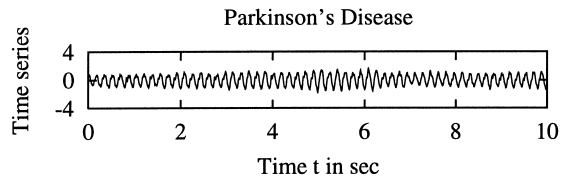
a)



b)



c)



point, as well, but with much weaker damping and larger rotation resulting in a lower inwards spiralling motion. In the case of Parkinson's disease in contrast, the deterministic parts of the dynamics are governed by limit cycles. The result of the analysis concerning the physiological tremor is in accordance with earlier studies on model equations for the dynamics of physiological tremor. In Ref. [10] it was shown that these dynamics can be explained by an AR(2) process driven by the uncorrelated firing of motoneurons in the arm muscles.

5. Conclusions

In this letter a method is presented which enables one to extract the underlying dynamic equations from observed data. It is possible to estimate the deterministic and the stochastic parts of this dynamics.

For validation, the method was applied to data from a chaotic electric circuit, first. The dynamic equations of this physical system were known and allowed a comparison of numerically determined and expected results.

Furthermore, the algorithm was applied to tremor data measured in normal subjects with physiological tremor, patients with essential tremor and patients with Parkinson's disease. The results allow a distinction of the different diseases. As this method allows

a better understanding of the underlying dynamics of tremors more insights into their pathophysiology are expected.

Concluding, it is expected that the presented method will have a huge field of possible applications in science and might become a new standard tool in the analysis of time series of nonlinear, complex systems.

Acknowledgements

The authors acknowledge helpful discussions with A. Kittel and Ch. Renner, Department of Physics, University of Oldenburg.

References

- [1] H. Haken, *Synergetik*, Springer, Berlin, 1990.
- [2] H. Haken, *Advanced Synergetics*, Springer, Berlin, 1983.
- [3] H. Haken, *Information and Self-Organization*, Springer, Berlin, 1988.
- [4] P. Hänggi, H. Thomas, *Phys. Rep.* 88 (1982) 207.
- [5] A.N. Kolmogorov, *Math. Ann.* 104 (1931) 415.
- [6] H. Risken, *The Fokker-Planck equation*, Springer, Berlin, 1984.
- [7] K. Itô, *Nagoya Mth. J* 1 (1950) 35.
- [8] S. Siegert, R. Friedrich, J. Peinke, *Phys. Lett. A* 243 (1998) 275.
- [9] M. Shinriki, M. Yamamoto, S. Mori, *Proc. IEEE* 69 (1981) 394.
- [10] J. Timmer, *Int. J. Bif. Chaos* 8 (1998) 1505.