Modeling of nonlinear Lévy processes by data analysis

Silke Siegert* and Rudolf Friedrich
Institute for Theoretical Physics 3, University of Stuttgart, Pfaffenwaldring 57, D-70550 Stuttgart, Germany
(Received 28 August 2000; published 25 September 2001)

The paper presents a method to analyze time series of nonlinear Lévy processes. The Lévy stability index as well as the nonlinear deterministic and stochastic parts of the dynamics together with their uncertainties can be calculated numerically. As last step of the analysis the membership of the investigated system to the regarded class of dynamical systems is validated. For demonstration the algorithm is applied to artificially created time series with different Lévy indices.

DOI: 10.1103/PhysRevE.64.041107 PACS number(s): 05.40.Fb, 05.45.−a, 05.10.Gg

I. INTRODUCTION

Normal diffusion under the influence of an external force field is often described by a Langevin equation for the d-dimensional stochastic variable $\mathbf{X}(t)$ (see, e.g., [1–3])

$$\mathbf{dX}(t)=\mathbf{g}(\mathbf{X}(t),t)dt+\mathbf{h}(\mathbf{X}(t),t)\mathbf{dW}(t). \tag{1}$$

The trajectory of the stochastic variable $\mathbf{X}(t)$ is determined by a deterministic part $\mathbf{g}$ and a stochastic part $\mathbf{h}$. $\mathbf{dW}$ stands for an infinitesimal Brownian motion, i.e., an infinitesimal Wiener process. The d-dimensional driving noise source $\mathbf{\Gamma}(t)$ of the Wiener process $\mathbf{dW}=\mathbf{\Gamma}(t)dt$ is a δ-function correlated Gaussian white noise with

$$\langle\Gamma_i(t)\rangle=0, \tag{2}$$

$$\langle\Gamma_i(t)\Gamma_j(t')\rangle=\delta_{ij}\delta(t-t'). \tag{3}$$

A lot of physical and biological systems can be described by models like Eq. (1) (see, e.g., [4,5]).

If the deterministic part $\mathbf{g}$ and the stochastic part $\mathbf{h}$ of Eq. (1) are not explicitly time dependent, one talks about a stationary process. In this case the probability density distribution $w(\mathbf{x},t)=w(\mathbf{x})$ in state space $\{\mathbf{x}\}$ is stationary, i.e. it does not change in time.

For such stationary stochastic processes that can be described by a Langevin equation (1) with $\mathbf{g}=\mathbf{g}(\mathbf{X}(t))$ and $\mathbf{h}=\mathbf{h}(\mathbf{X}(t))$ a data-driven method to determine the deterministic and stochastic parts of the dynamics by data analysis directly from measured time series was proposed in Refs. [6,7].

The results of an application of the method to experimental data sets originating from physical, technical, and medical research was shown in Refs. [8–10].

The method can be extended to an application on nonstationary systems by a moving window technique. Then, the dynamics is assumed to be quasistationary within each window.

In this paper, the class of Langevin systems (1) will be extended to the bigger class of Langevin-like systems where the Gaussian white noise function is replaced by the more general Lévy noise (see, e.g., [11–13]). Stochastic Lévy processes constitute an interesting generalization of normal diffusion processes (see, e.g., [14–16]). The typical length of Lévy flights grows according to

$$\langle|x|\rangle\sim t^{1/\alpha},\quad \alpha\in(0,2], \tag{4}$$

where $\alpha=2$ stands for the behavior of normal diffusion. This property has made Lévy flights natural candidates for the description of enhanced diffusion.

The distribution of Lévy noise has a long-range algebraic tail corresponding to large but infrequent steps, so-called rare events. Therefore, rare events become more important in Lévy flights than in Brownian motion. The superdiffusive characteristics of Lévy flights have recently been used to model a broad variety of physical processes. Some examples are the description of anomalous transport in one dimension with absorbing boundary [16], the modeling of anomalous diffusion at liquid surfaces (so-called bulk-mediated surface diffusion), where the molecules execute Lévy walks on the surface [17], and the application of this modeling in the cases of porous glasses [18] and eye lenses [19]. Enhanced diffusion has been observed in systems of polymerlike breakable micelles [20]. Turbulence has been investigated under the point of view of Lévy flights [21]. Even the wandering of albatrosses has been modeled by the theory of Lévy flights [22].

Among the different theoretic frameworks connected with the description of anomalous diffusion are continuous random walk schemes [23,24], fractional diffusion equations [25,15], and generalized Langevin and Fokker-Planck equations [14].

In the following the theoretical model of a generalized Langevin equation as generalization of the Langevin equation for normal diffusion processes (1) will be used. For dynamic systems that are describable by such an evolution equation an analysis method will be presented that provides us with a direct data-driven tool to formulate model equations for the dynamics of the system.

II. CONSIDERED SYSTEMS

Instead of Eq. (1) the following nonlinear Langevin-like differential equation for a stochastic vector $\mathbf{X}(t)$ is assumed to describe the dynamics of the investigated systems (see, e.g., [11–13,15]):

*Email address: silke@theo3.physik.uni-stuttgart.de

$$\mathbf{dX}(t)=\mathbf{g}(\mathbf{X}(t),t)+\mathbf{h}(\mathbf{X}(t),t)\mathbf{dL}_{\alpha}^{(\gamma,\beta,\mu)}. \tag{5}$$

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

$$(\mathbf{dL}_{\alpha}^{(\gamma,\beta,\mu)})_i=(\mathbf{f}_{\alpha}^{(\gamma,\beta,\mu)})_i(t)dt, \tag{6}$$

where the d stochastically independently chosen components $(\mathbf{f}_{\alpha}^{(\gamma,\beta,\mu)})_i(t)$ of the Lévy noise creating such a motion are characterized by the four parameters: Lévy stability $\alpha \in (0,2]$, scale parameter $\gamma \geqslant 0$, skewness $\beta \in [-1,1]$, and center $\mu \in \mathbb{R}$. For $\alpha=2$, $\gamma=0.5$, $\beta=0$, and $\mu=0$ Eq. (5) is equal to the original Langevin equation (1). For decreasing α the larger deviations of the realization become larger and more frequent. On the left side of Fig. 4 realizations of Lévy noise are shown for different values of α.

For simplicity of the model

$$\mathbf{dX}(t)=\mathbf{g}(\mathbf{X}(t),t)+\tilde{\mathbf{h}}(\mathbf{X}(t),t)\widetilde{\mathbf{dL}}_{\alpha}^{(\gamma,\beta,\mu)} \tag{7}$$

with

$$\tilde{h}_{ij}(\mathbf{X}(t),t)=\tilde{h}_{ii}(\mathbf{X}(t),t)\delta_{ij}, \quad \tilde{h}_{ii}>0, \tag{8}$$

$$\widetilde{\mathbf{dL}}_{\alpha}^{(\gamma,\beta,\mu)}=\tilde{\mathbf{f}}_{\alpha}^{(\gamma,\beta,\mu)}dt \tag{9}$$

is considered instead of Eq. (5). Now, the d components $(\tilde{\mathbf{f}}_{\alpha}^{(\gamma,\beta,\mu)})_i$ of the introduced Lévy-noise creating the infinitesimal Lévy-motions $(\widetilde{dL}_{\alpha}^{(\gamma,\beta,\mu)})_i$ are no longer necessarily stochastically independent.

In general, Lévy noise is not defined by its probability density, but by its characteristic function. The probability density distribution exists and is continuous but, with some exceptions, it is not known in closed form. The characteristic function of each Lévy noise component, i.e. the Fourier transform of the probability density distribution $p_{\alpha}^{(\gamma,\beta,\mu)}$ of each component, has the following form:

$$\begin{aligned}\langle e^{iqx}\rangle &= \int_{-\infty}^{\infty} p_{\alpha}^{(\gamma,\beta,\mu)}(x)e^{iqx}\,dx \\ &= \begin{cases} \exp\left\{-\gamma|q|^{\alpha}\left(1-i\beta(\operatorname{sgn} q)\tan\dfrac{\pi\alpha}{2}\right)+i\mu q\right\} \\ \quad \text{if} \quad \alpha\neq 1, \\ \exp\left\{-\gamma|q|\left(1+i\beta\dfrac{2}{\pi}(\operatorname{sgn} q)\ln|q|\right)+i\mu q\right\} \\ \quad \text{if} \quad \alpha=1. \end{cases}\end{aligned} \tag{10}$$

It is the most general form of a characteristic function of a stable process. For $(\beta=0,\mu=0)$ the probability density distribution is given by

$$p_{\alpha}^{(\gamma,\beta=0,\mu=0)}(x)=\frac{1}{\pi}\int_0^{\infty}\exp(-\gamma|q|^{\alpha})\cos(qx)\,dq. \tag{11}$$

For the following considerations, $\gamma=1$, $\mu=0$ is set without loss of generality for the model equations (5) and (7). Besides, $\beta=0$ is chosen. In this case, a series expansion of Eq. (11) valid for large arguments $(|x| \gg 0)$ is given by

$$p_{\alpha}^{(\gamma=1,\beta=0,\mu=0)}(|x|)=-\frac{1}{\pi}\sum_{k=1}^{n}\frac{(-1)^k}{k!}\frac{\Gamma(\alpha k+1)}{|x|^{\alpha k+1}}\sin\left[\frac{k\pi\alpha}{2}\right] + R(|x|), \tag{12}$$

where $\Gamma(x)$ is the Euler Γ function and

$$R(|x|)=\mathcal{O}(|x|^{-\alpha(n+1)-1}). \tag{13}$$

From this expansion, an asymptotic approximation of a stable distribution of index α for large values of $|x|$ can be found as

$$p_{\alpha}^{(\gamma=1,\beta=0,\mu=0)}(|x|)\sim\frac{\Gamma(1+\alpha)\sin(\pi\alpha/2)}{\pi|x|^{1+\alpha}}\sim|x|^{-(1+\alpha)}. \tag{14}$$

The asymptotic behavior for large values of $|x|$ is a power-law behavior. It results in a divergence of all moments $\langle |x|^n \rangle$ with $n \geqslant \alpha$ when $\alpha<2$. In particular, all Lévy stable processes with $\alpha<2$ have infinite variance.

In the following, a method to determine the Lévy stability index α, when $\alpha \in (1,2]$ and the deterministic and stochastic parts of a Langevin-like equation (7) directly from given time series will be presented. For the analysis the process has to be stationary, i.e., deterministic and stochastic parts are not explicitly time dependent. If instationary systems are investigated, a moving window technique has to be applied, i.e., the system is assumed to be stationary within one window and the time dependence of the functionals is estimated later on by taking together the results of the overlapping windows.

These considerations together with the considerations about the Lévy parameters above lead to the following class of model equations for the investigated considered dynamic systems:

$$\mathbf{dX}(t)=\mathbf{g}(\mathbf{X}(t))+\tilde{\mathbf{h}}(\mathbf{X}(t))\widetilde{\mathbf{dL}}_{\alpha}^{(\gamma=1,\beta=0,\mu=0)}, \quad \alpha\in(1,2], \quad h_{ij}=h_{ii}\delta_{ij}. \tag{15}$$

For small τ the evolution equation (15) is integrated and iterated in an analogous way to the Euler method as:

$$\mathbf{X}(t+\tau)\approx\mathbf{X}(t)+\mathbf{g}(\mathbf{X}(t))\tau+\tilde{\mathbf{h}}(\mathbf{X}(t))\tau^{1/\alpha}\mathbf{f}_{\alpha}^{(\gamma=1,\beta=0,\mu=0)}(t), \quad \tau\ll 1. \tag{16}$$

This iteration may be understood as definition of the differential equation (15) and will form the basis of the procedure described in the following.

III. A METHOD FOR ANALYZING LÉVY-STOCHASTIC SYSTEMS

For the extracting of deterministic and stochastic parts from fluctuating data, first, it is assumed that the underlying dynamics can be described by an iteration like Eq. (16). In the end, after all analysis results are found, this assumption can be verified.

Because of $\beta=\mu=0$ the function $\mathbf{g}(\mathbf{x})$ in state space $\{\mathbf{x}\}$ (i.e. the space $\{\mathbf{x}\}$ of all values $\mathbf{x}$ that can be taken by the stochastic variable $\mathbf{X}(t)$) can be expressed as conditional average according to

$$\mathbf{g}(\mathbf{x})=\lim_{\tau\to 0}\frac{1}{\tau}\langle \mathbf{X}(t+\tau)-\mathbf{x}\rangle|_{\mathbf{X}(t)=\mathbf{x}}=\lim_{\tau\to 0}\frac{1}{\tau}\mathbf{T}^{(1)}(\mathbf{x},\tau) \tag{17}$$

with

$$\mathbf{T}^{(1)}(\mathbf{x},\tau):=\langle \mathbf{X}(t+\tau)-\mathbf{x}\rangle|_{\mathbf{X}(t)=\mathbf{x}}. \tag{18}$$

If stationarity is given, what has been assumed, the deterministic and stochastic functions $\mathbf{g}$ and $\tilde{\mathbf{h}}$ are not explicitly time dependent. So, the ensemble average $\mathbf{T}^{(1)}$ can be estimated as conditional temporal average over the whole time series $\mathbf{X}(t_n)$. For this, the state space of the process is discretized. The condition $\mathbf{X}(t)=\mathbf{x}$ is fulfilled if only pairs $\mathbf{X}(t_n+\tau),\mathbf{X}(t_n)$ with $\mathbf{X}(t_n)=\mathbf{x}$ within some limits $\Delta\mathbf{x}$, i.e. with $\mathbf{X}(t_n)$ and $\mathbf{x}$ lying within the same bin of the discretized state space, are taken into consideration. In the following this is expressed by the condition $\mathbf{X}(t_n)=\mathbf{x}\pm\Delta\mathbf{x}$. The conditional average $\mathbf{T}^{(1)}$ has to be calculated for every $\mathbf{x}$ separately. N shall be the number of data pairs for a certain $\mathbf{x}$ that fulfill the condition. Then, $\mathbf{T}^{(1)}$ can be estimated as

$$\mathbf{T}^{(1)}(\mathbf{x},\tau)\overset{N\text{large}}{\approx}\mathbf{T}_E^{(1)(N)}(\mathbf{x},\tau):=\frac{1}{N}\sum_{n=1}^{N}(\mathbf{X}(t_n+\tau)-\mathbf{X}(t_n))\Big|_{\mathbf{X}(t_n)=\mathbf{x}\pm\Delta\mathbf{x}}. \tag{19}$$

The subscript $_E$ will be added whenever an estimation value for an observable is introduced. Generally, N is different for every point $\mathbf{x}$ and has to be large for a good result (see uncertainty discussion in next section). To be reminded that N has an influence on the estimated value even if it is not treated as direct functional dependence, the index (N) is added as superscript to the estimation value.

If $\mathbf{T}_E^{(1)(N)}$ is calculated for different (but small) values of τ like e.g. $\Delta t_{samp.}$, $2\Delta t_{samp.}$, $3\Delta t_{samp.}$, ..., where $\Delta t_{samp.}$ is the sampling time of the time series, the limit $\tau\to 0$ of $\mathbf{T}^{(1)}/\tau$ according to relationship (17) can be found by extrapolation. By this way, an estimation $\mathbf{g}_E^{(N)}(\mathbf{x})$ for the deterministic part $\mathbf{g}(\mathbf{x})$ of the dynamics can be calculated.

To estimate the stochastic part $\tilde{\mathbf{h}}(\mathbf{x})$, the Lévy index α has to be determined, first. Therefore, the following conditional average is considered and approximated for small τ as

$$\begin{aligned}\langle|X_i(t+\tau)-x_i-g_i(\mathbf{x})\tau|\rangle|_{\mathbf{X}(t)=\mathbf{x}}&=T_i^{(2)}(\mathbf{x},\tau)\approx\tilde{h}_{ii}(\mathbf{x})\tau^{1/\alpha}\int_{-\infty}^{\infty}|x'|p_\alpha^{(\gamma=1,\beta=0,\mu=0)}(x')dx'\\&=\tilde{h}_{ii}(\mathbf{x})\tau^{1/\alpha}F(\alpha),\quad \tau\ll 1,\end{aligned} \tag{20}$$

respectively,

$$\ln[T_i^{(2)}(\mathbf{x},\tau)]\approx\ln[\tilde{h}_{ii}(\mathbf{x})F(\alpha)]+\frac{1}{\alpha}\ln(\tau),\quad \tau\ll 1, \tag{21}$$

with

$$T_i^{(2)}(\mathbf{x},\tau):=\langle|X_i(t+\tau)-x_i-g_i(\mathbf{x})\tau|\rangle|_{\mathbf{X}(t)=\mathbf{x}}, \tag{22}$$

$$F(\alpha):=\int_{-\infty}^{\infty}|x'|p_\alpha^{(\gamma=1,\beta=0,\mu=0)}(x')dx'. \tag{23}$$

For $\mathbf{g}(\mathbf{x})$ being known in form of the estimation $\mathbf{g}_E^{(N)}(\mathbf{x})$ in the meantime, the conditional averages $T_i^{(2)}(\mathbf{x},\tau)$ can be estimated as

$$\begin{aligned}T_i^{(2)}(\mathbf{x},\tau)&\approx T_{iE}^{(2)(N)}(\mathbf{x},\tau)\\&:=\frac{1}{N}\sum_{n=1}^{N}|X_i(t_n+\tau)-X_i(t_n)-g_{iE}^{(N)}(\mathbf{x})\tau|\Big|_{\mathbf{X}(t_n)=\mathbf{x}\pm\Delta\mathbf{x}}.\end{aligned} \tag{24}$$

The proceeding is analogous to the calculations described above.

When $T_i^{(2)}$ is calculated for several small values of τ (e.g. Δt_{samp}, $2\Delta t_{samp}$, $2^2\Delta t_{samp}$, ...) and is plotted for fixed $\mathbf{x}$ and i in a ln-ln-plot over τ, α can be determined as inverse slope of a fitted straight line. In theory, for each $\mathbf{x}$ and i the same value for α should be received by this method. In praxis, errors because of finite time series, finite discretization of state space, extrapolation, and measurement (see Sec. IV) can be minimized by taking the mean of all values for α, that have been calculated for the different $\mathbf{x}$ and i. This average value α_E is used as an estimate for the Lévy index. The standard deviation of this distribution can be taken as uncertainty $\Delta\alpha_E$.

Now, expression $F(\alpha)$ can be estimated numerically according to Eq. (23) by simulating a Lévy motion with ($\alpha=\alpha_E,\gamma=1,\beta=0,\mu=0$) and taking the average of the absolute value of the realization. The realization $f_\alpha^{(\gamma=1,\beta=0,\mu=0)}$ of the Lévy noise can be constructed as follows [26,27]: A realization r of a uniformly distributed random variable in the interval $[-\pi/2,\pi/2]$ is taken and, independently, a realization v of an exponential random variable with mean 1. Then

$$f_\alpha^{(\gamma=1,\beta=0,\mu=0)}=\frac{\sin(\alpha r)}{(\cos r)^{1/\alpha}}\left(\frac{\cos[(1-\alpha)r]}{v}\right)^{(1-\alpha)/\alpha}. \tag{25}$$

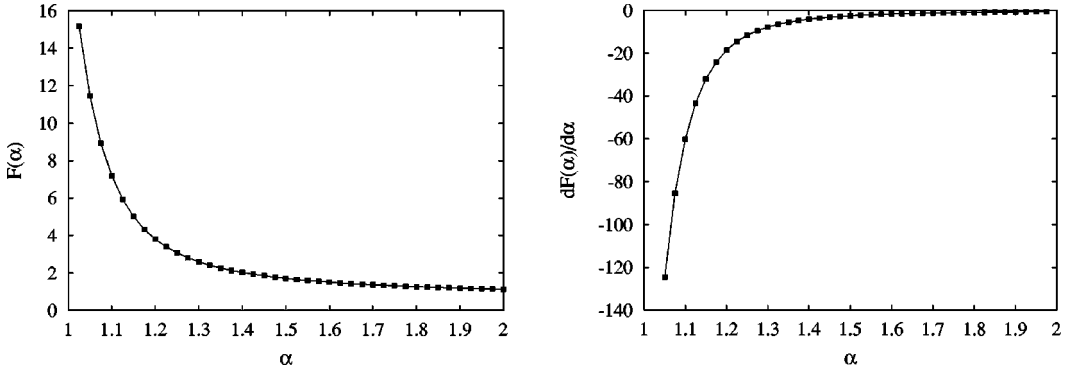

FIG. 1. Numerically determined integral $F(\alpha)$ and the derivative $F'(\alpha)=dF(\alpha)/d\alpha$ over α.

The numerical result for $F(\alpha)$ in the range $\alpha\in(1.0,2.0]$ is presented on the left side of Fig. 1.

The stochastic part $\tilde{\mathbf{h}}(\mathbf{x})$ is given by

$$\tilde{h}_{ii}(\mathbf{x})=\lim_{\tau\to 0}\frac{\langle|X_i(t+\tau)-x_i-g_i(\mathbf{x})\tau|\rangle|_{\mathbf{X}(t)=\mathbf{x}}}{\tau^{1/\alpha}F(\alpha)}$$

$$=\lim_{\tau\to 0}\frac{T_i^{(2)}(\mathbf{x},\tau)}{\tau^{1/\alpha}F(\alpha)}. \tag{26}$$

It can be estimated as

$$\tilde{h}_{iiE}^{(N,\alpha_E)}(\mathbf{x})=\lim_{\tau\to 0}\frac{T_{iE}^{(2)(N)}(\mathbf{x},\tau)}{\tau^{1/\alpha_E}F(\alpha_E)}. \tag{27}$$

Additionally to N, the estimated Lévy index α_E has been added as superscript to remind on a possible uncertainty propagation caused by this estimation value. The procedure concerning the calculation of the limit $\tau\to 0$ is analogous to the determination of the deterministic part.

IV. UNCERTAINTIES OF THE RESULTS

Up to this point of the analysis just the uncertainty of the Lévy index α has been concerned. But for experimentally as well as for numerically set up values error bars are of great importance for a correct interpretation and further application of the results. In addition to this quite general justification for uncertainty discussions, for the presented algorithm error bars have an additional importance. The necessary extrapolations to $\tau=0$ can be calculated with greater accuracy if explicate uncertainties for the τ-dependent values are known compared with an extrapolation under the assumption of unit standard deviations. The analysis results calculated so far for the Lévy index α, the deterministic part $\mathbf{g}(\mathbf{x})$ and the stochastic part $\tilde{\mathbf{h}}(\mathbf{x})$ provide us with first (already very good) estimations for the real observables. With the help of these values uncertainties can be calculated as will be described below. Afterwards the extrapolations to $\tau=0$ can be determined again, this time under the use of real error bars instead of unit standard deviations. So, new estimations for α, $\mathbf{g}(\mathbf{x})$ and $\tilde{\mathbf{h}}(\mathbf{x})$ can be calculated. They do not differ very much from the first estimations, anyhow. This procedure may be run a few times, calculating new error bars under the use of the estimations, determining new estimations with the help of the error bars. One will recognize that the procedure is convergent after at most one or two runs. In the following, the analytic expressions for the uncertainties of interest will be presented.

The term $\mathbf{T}_E^{(1)(N)}(\mathbf{x},\tau)$ is calculated as sum of N independently identically Lévy distributed stochastic variables. Therefore, $\mathbf{T}_E^{(1)(N)}(\mathbf{x},\tau)$ itself is a stochastic variable, which is also Lévy distributed.

$$\frac{1}{\tau}\mathbf{T}_E^{(1)(N)}(\mathbf{x},\tau)=\frac{1}{\tau}\frac{1}{N}\sum_{n=1}^{N}[\mathbf{X}(t_n+\tau)-\mathbf{X}(t_n)]|_{\mathbf{X}(t_n)=\mathbf{x}\pm\Delta\mathbf{x}}$$

$$\approx\mathbf{g}(\mathbf{x})+\tilde{\mathbf{h}}(\mathbf{x})\frac{1}{\tau^{1-1/\alpha}}\mathbf{f}_\alpha^{(\gamma=1/N^{\alpha-1},\beta=0,\mu=0)}$$

$$=\mathbf{g}(\mathbf{x})+\tilde{\mathbf{h}}(\mathbf{x})\frac{1}{(\tau N)^{1-1/\alpha}}\mathbf{f}_\alpha^{(\gamma=1,\beta=0,\mu=0)}. \tag{28}$$

As uncertainty of a Lévy distributed stochastic variable $Z=f_\alpha^{(\gamma=1,\beta=0,\mu=0)}$ the width $\Delta_\alpha^{(70)}$ is introduced with

$$\int_{-\Delta_\alpha^{(70)}}^{\Delta_\alpha^{(70)}}p(z)dz\geqslant 0.7. \tag{29}$$

$\Delta_\alpha^{(70)}$ can be calculated numerically and is illustrated in Fig. 2.

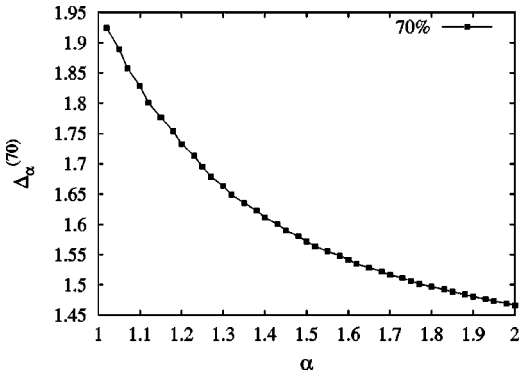

FIG. 2. Numerically determined uncertainty $\Delta_\alpha^{(70)}$, defined by Eq. (29), over α.

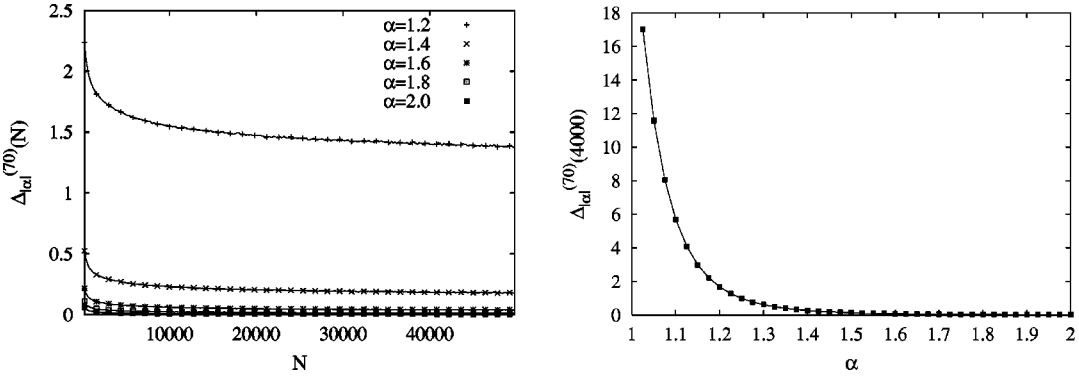

FIG. 3. Numerically determined uncertainty $\Delta_{|\alpha|}^{(70)}$, defined by Eq. (32), over N for fixed α (left side) and over α for fixed $N=4000$ (right side).

With the help of definition (29) the uncertainty of $1/\tau \mathbf{T}_E^{(1)(N)}(\mathbf{x},\tau)$ can be expressed as

$$\frac{1}{\tau}\Delta \mathbf{T}_E^{(1)(N)}(\mathbf{x},\tau)=\tilde{\mathbf{h}}_E^{(N)}(\mathbf{x})\frac{1}{(\tau N)^{1-1/\alpha_E}}\Delta_{\alpha_E}^{(70)}. \tag{30}$$

The results together with their uncertainty limits have a confidence of 70%. This width of the confidence interval is used in the whole contribution whenever a standard deviation is not applicable. Now, the confidence of $\mathbf{g}_E^{(N)}(\mathbf{x})$ can be determined by extrapolation of $1/\tau \mathbf{T}_E^{(1)(N)}(\mathbf{x},\tau)$ to $\tau=0$ under consideration of the error bars.

The influence of the uncertainty $\Delta\mathbf{x}$ of $\mathbf{x}$ will not be considered explicitly in the following. It is taken into account as uncertainty in phase space and has the value of the discretization width. If differentiable functions for $\mathbf{g}(\mathbf{x})$ and $\tilde{\mathbf{h}}(\mathbf{x})$ are assumed, the estimated values in every bin are always within the interval from the minimum to the maximum of the actual, correct function within this bin. So, the estimated values for deterministic and stochastic parts in each bin will represent the correct value of the observables at least somewhere within the considered bin.

The uncertainty of $T_{iE}^{(2)(N)}(\mathbf{x},\tau)$ can be found in an equivalent way to the determination of the uncertainty of $\mathbf{T}_E^{(1)(N)}$. The only difficulty in this case is the fact that $T_{iE}^{(2)(N)}$ is not a sum of independently identically Lévy distributed stochastic variables, but the sum of absolute values of such variables. Therefore, $T_{iE}^{(2)(N)}$ itself is of course again a stochastic variable, but in general not Lévy distributed anymore.

$$\begin{aligned} T_{iE}^{(2)(N)}(\mathbf{x},\tau) &= \frac{1}{N}\sum_{n=1}^{N}\left|[X_i(t_n+\tau)-X_i(t_n)-g_i(\mathbf{x})\tau]\right|\Big|_{\mathbf{X}(t_n)=\mathbf{x}\pm\Delta\mathbf{x}} \\ &\overset{\tau\ll 1}{\approx} \tilde{h}_{ii}(\mathbf{x})\tau^{1/\alpha}\frac{1}{N}\sum_{n=1}^{N}\left|f_\alpha^{(\gamma=1,\beta=0,\mu=0)}(t_n)\right| \\ &\overset{N\to\infty}{\to} \tilde{h}_{ii}(\mathbf{x})\tau^{1/\alpha}F(\alpha). \end{aligned} \tag{31}$$

The uncertainty of $T_{iE}^{(2)(N)}$ can be expressed with the help of a new defined width $\Delta_{|\alpha|}^{(70)}(N)$: As uncertainty of a stochastic variable $Z=1/N\Sigma_{n=1}^{N}|f_\alpha^{(\gamma=1,\beta=0,\mu=0)}|$ the width $\Delta_{|\alpha|}^{(70)}(N)$ is defined by

$$\int_{F(\alpha)-\Delta_{|\alpha|}^{(70)}(N)}^{F(\alpha)+\Delta_{|\alpha|}^{(70)}(N)} p(z)dz \geqslant 0.7. \tag{32}$$

$\Delta_{|\alpha|}^{(70)}(N)$ can be determined numerically by simulating Lévy processes as described above, calculating the stochastic variable Z with fixed N for statistically representative times and estimating $\Delta_{|\alpha|}^{(70)}(N)$ for the distribution. On the left side of Fig. 3 the results for $\Delta_{|\alpha|}^{(70)}(N)$ are plotted over N for different values of α. The right side of the figure illustrates the dependence of the width on α for fixed $N=4000$.

Now, the uncertainty of $T_{iE}^{(2)}$ can be calculated as

$$\Delta(T_{iE}^{(2)(N)}(\mathbf{x},\tau))=\tilde{h}_{iiE}^{(N,\alpha_E)}(\mathbf{x})\tau^{1/\alpha_E}\Delta_{|\alpha_E|}^{(70)}(N), \tag{33}$$

respectively,

$$\begin{aligned} \Delta(\ln[T_{iE}^{(2)(N)}(\mathbf{x},\tau)]) &= \frac{1}{T_{iE}^{(2)(N)}(\mathbf{x},\tau)}\Delta T_{iE}^{(2)(N)}(\mathbf{x},\tau) \\ &= \frac{1}{T_{iE}^{(2)(N)}(\mathbf{x},\tau)}\tilde{h}_{iiE}^{(N,\alpha_E)}(\mathbf{x})\tau^{1/\alpha_E}\Delta_{|\alpha_E|}^{(70)}(N). \end{aligned} \tag{34}$$

These uncertainty intervals can be used in the determination of the Lévy index α when a straight line is fitted to the data points. The determination of the error of α_E itself has already been discussed above. It is taken as the standard deviation of all single values determined for different $\mathbf{x}$ and i.

As the last step, the uncertainty of $\tilde{h}_{iiE}$ has to be discussed.

$$\tilde{h}_{iiE}^{(N,\alpha_E)}(\mathbf{x})=\lim_{\tau\to 0}\frac{T_{iE}^{(2)(N)}(\mathbf{x},\tau)}{\tau^{1/\alpha_E}F(\alpha_E)}=\lim_{\tau\to 0}T_{iE}^{(3)(N,\alpha_E)}(\mathbf{x},\tau) \tag{35}$$

with

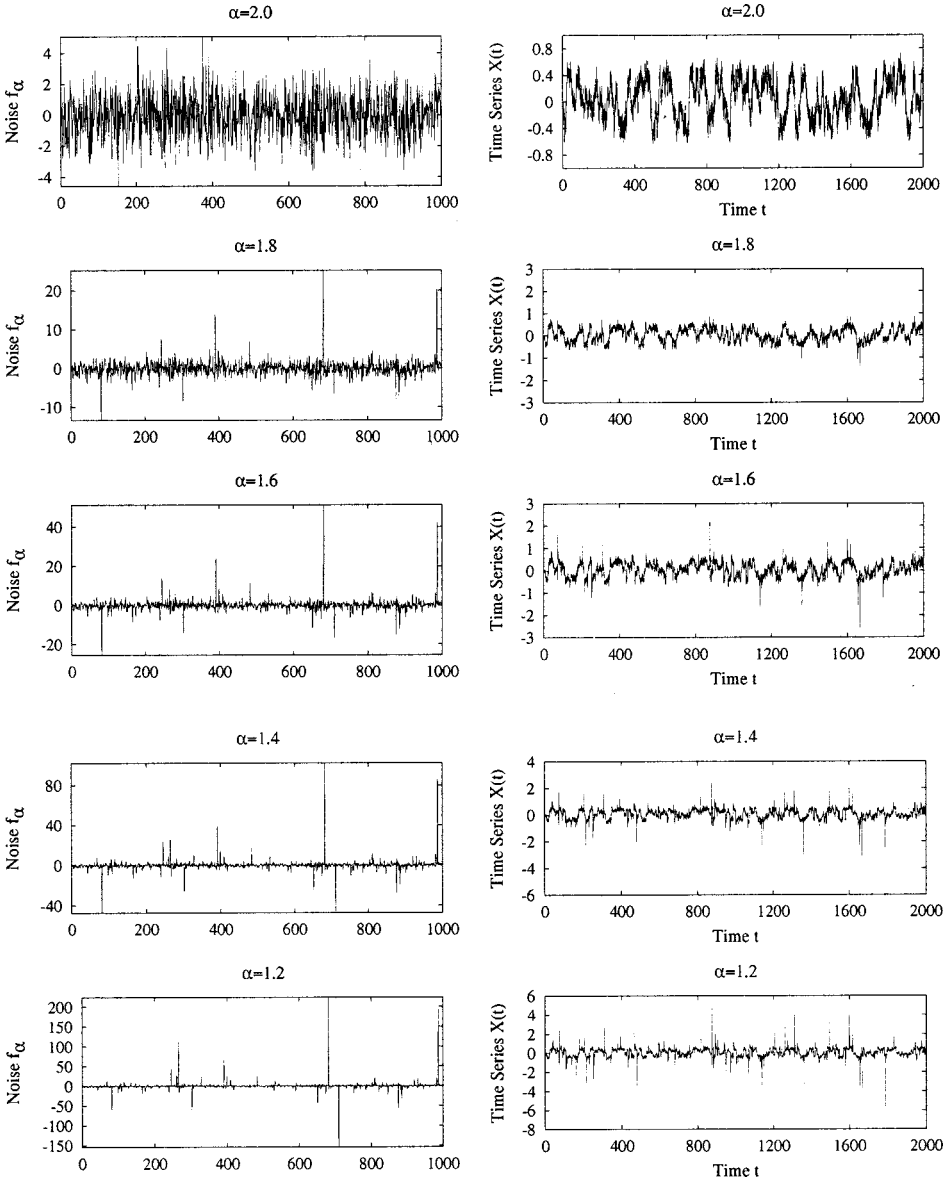

FIG. 4. Realizations of Lévy noise for different values of the Lévy index α and subsequences of the affiliated integrated nonlinear Lévy motion according to the relations (40)–(42).

$$T_{iE}^{(3)(N,\alpha_E)}(\mathbf{x},\tau) := \frac{T_{iE}^{(2)(N)}(\mathbf{x},\tau)}{\tau^{1/\alpha_E} F(\alpha_E)}. \tag{36}$$

Errors of this estimation are caused by the uncertainty of $T_{iE}^{(2)(N)}$, which is primarily originated in the finite number N of relevant data points for every bin $\mathbf{x}[\mathbf{X}(t_n) = \mathbf{x} \pm \Delta\mathbf{x} \ \ \forall n]$, and by the uncertainty of α_E which has to be inserted in the formulas (35) and (36). The total uncertainty of $T_{iE}^{(3)(N,\alpha_E)}$ can be expressed under consideration of the law of error propagation as

$$\begin{aligned}
&\Delta\left(T_{iE}^{(3)(N,\alpha_E)}(\mathbf{x},\tau)\right) \\
&\quad = \frac{\Delta T_{iE}^{(2)(N)}(\mathbf{x},\tau)}{\tau^{1/\alpha_E} F(\alpha_E)} + \frac{T_{iiE}^{(2)(N)}(\mathbf{x},\tau)}{\tau^{2/\alpha_E} F^2(\alpha_E)} \\
&\qquad \times \left(\left| F'(\alpha_E)\tau^{1/\alpha_E} - \tau^{1/\alpha_E} \ln\tau \frac{F(\alpha_E)}{\alpha_E^2} \right| \right) \Delta\alpha_E \\
&\quad = \frac{\tilde{h}_{iiE}^{(N,\alpha_E)}(\mathbf{x}) \Delta_{|\alpha_E|}^{\{70\}}(N)}{F(\alpha_E)} + \frac{T_{iE}^{(2)(N)}(\mathbf{x},\tau)}{\tau^{1/\alpha_E} F(\alpha_E)} \\
&\qquad \times \left(\frac{F'(\alpha_E)}{F(\alpha_E)} - \frac{\ln\tau}{\alpha_E^2} \right) \Delta\alpha_E .
\end{aligned} \tag{37}$$

The numerical results for the derivative $F'(\alpha)$ are illustrated on the right side of Fig. 1.

V. VALIDATION OF THE ASSUMPTIONS

The only assumption concerning the investigated system that has been made is the pure describability of the dynamics by an evolution equation like Eq. (16). No further ansatz has been taken into account, no further knowledge about the system of interest is necessary. Now, after the application of the algorithm to the data set, when the deterministic and stochastic parts as well as the Lévy index are known, the assumption

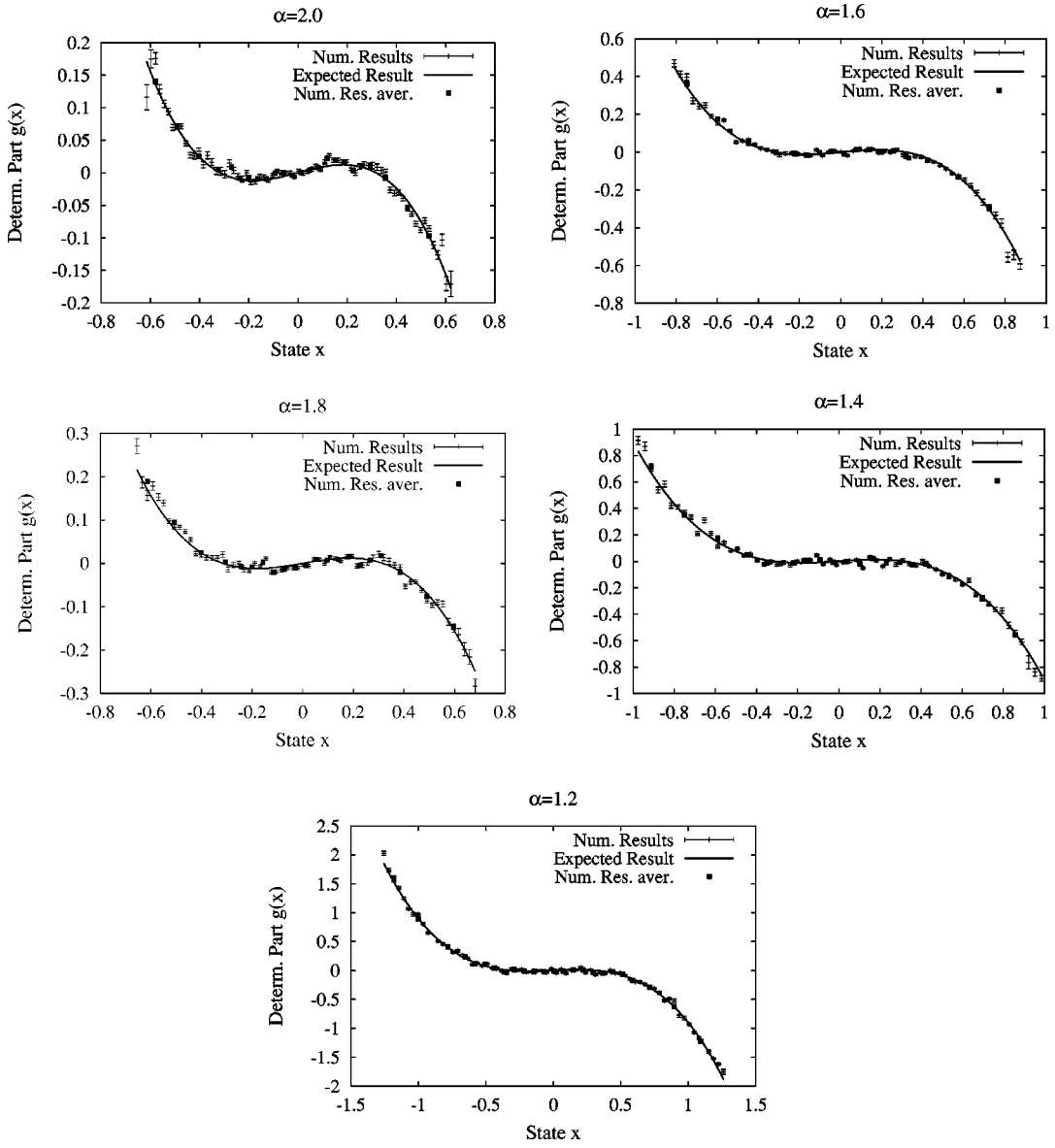

FIG. 5. Numerically determined values for the deterministic function $g(x)$ of system (40)–(42) in state space $\{x\}$ for different Lévy indices α. The values were calculated directly by data analysis of the time series shown in subsequences in Fig. 4, were smoothed, and were compared with the theoretical curve.

of the system belonging to this class of Langevin-like systems can be tested. Therefore, a new numerically calculated time series $\mathbf{Y}(t)$ can be integrated according to

$$Y_i(t+\Delta t_{sim.}) = g_{iE}(\mathbf{Y}(t))\Delta t_{sim} + \tilde{h}_{iiE}(\mathbf{Y}(t))(\Delta t_{sim})^{(1/\alpha_E)} f^{(\gamma=1,\beta=0,\mu=0)}_{(\alpha_E)}(t). \tag{38}$$

Δt_{sim} should be in the range of the internal dynamical time scale. If the dynamical structure is known it is possible to choose a suitable integration step. Besides, the relation

$$\Delta t_{samp} = n\Delta t_{sim} \tag{39}$$

should be fulfilled with large n and Δt_{samp} being the sampling time of the measured and investigated time series $\mathbf{X}(t)$.

The needed values for the deterministic and stochastic functions can be found by interpolation within that part of the state space where numerical results exist. In the outer part of the phase space, for large deviations of the time series, the numerical results can be extrapolated. According to the law of rare events the extrapolation is done in a linear way.

If the investigated time series $\mathbf{X}(t)$ is describable by an iteration like Eq. (16) then all statistical qualities of both series $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ should be equal within the range of the determined uncertainties. As one example, the conditional probability densities $p(x_1, t+\Delta t_{samp}|x_0, t)$ can be calculated for every x_0 for both time series and the functions can be compared. If there is a good agreement of the distributions

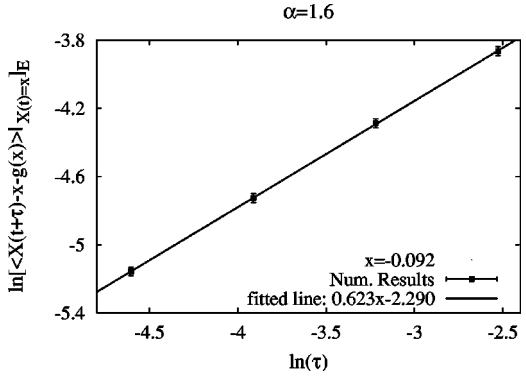

FIG. 6. Logarithm of the numerically calculated conditional average $1/N\Sigma_{n=1}^{N}[X(t_n+\tau)-X(t_n)-g(x)\tau]|_{X(t)=x\pm\Delta x}$ of the stochastic variable $X(t)$ over the logarithm of the time difference τ with $x=0.092$. The straight line was fitted to the numerically calculated values represented by the points with error bars.

for all x_0 the investigated time series $\mathbf{X}(t)$ belongs to the class of Langevin-like equations and it was justified to make the assumption of the algorithm.

VI. APPLICATION OF THE ALGORITHM

Results of an application of the presented algorithm to artificially created data sets will be shown in this section. As example, the dynamical system of a one-dimensional Pitchfork bifurcation with dynamical Lévy noise is used,

$$dX(t)=g(X(t))+h(X(t))dL_{\alpha}^{(\gamma=1,\beta=0,\mu=0)}(t) \tag{40}$$

with

$$g(x)=0.1x-x^3, \tag{41}$$

$$h(x)=\frac{1}{1000x^2+2000\sqrt{0.1}x+120}+\frac{1}{1000x^2-2000\sqrt{0.1}x+120}+0.05.$$

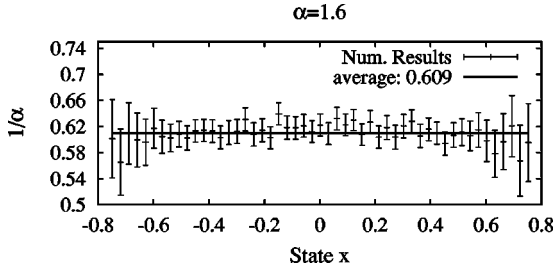

FIG. 7. Numerically determined values for $1/\alpha$, α being the Lévy index, over state x. The calculated values are represented by points with error bars, the straight line represents the average of all values.

TABLE I. Theoretical values and numerically determined values with uncertainties for different Lévy indices α and their inverse $1/\alpha$.

α	$1/\alpha$	α num.	$1/\alpha$ num.
2.0	0.50	2.01 ± 0.03	0.50 ± 0.01
1.8	0.56	1.82 ± 0.03	0.55 ± 0.01
1.6	0.63	1.64 ± 0.04	0.61 ± 0.01
1.4	0.71	1.45 ± 0.06	0.69 ± 0.03
1.2	0.83	1.28 ± 0.09	0.78 ± 0.06

$h(x)$ has been chosen in this way to produce a positive function with two maxima and constant asymptotic behavior for $|x|\to\infty$. The differential equation has been integrated numerically according to

$$X(t+\Delta t_{sim})=X(t)+g(X(t))\Delta t_{sim}+h(X(t))(\Delta t_{sim})^{1/\alpha}f_{\alpha}^{(\gamma=1,\beta=0,\mu=0)}(t). \tag{42}$$

Δt_{sim} has been chosen as 0.001 time units, but just every tenth data point of the sample path has been stored for further analysis. This corresponds to a sampling time step of $\Delta t_{samp}=0.01$ time units. For α the values 2.0,1.8,1.6,1.4,1.2 have been taken. Figure 4 illustrates Lévy noise realizations for these Lévy indices and subsequences of the corresponding integrated motions according to the relations (40)–(42).

For each of the time series the deterministic part of the underlying dynamics has been determined according to relation (17). The results are shown in Fig. 5. The numerically determined values are represented by points with error bars, which are quite small. The numerical results have been averaged to eliminate smaller fluctuations, the results are given by circles. The solid line is the expected curve according to Eq. (41).

If $g(x)$ is known, α can be determined. As an example for the proceeding that has been described in detail above, the ln-ln-plot and the fitted straight line for $\alpha=1.6$ and $x=-0.092$ are shown in Fig. 6. Figure 7 shows the distribution of all numerically calculated results for α for different x in the case $\alpha=1.6$ and the affiliated mean value.

In Table I the numerically determined values for $1/\alpha$ and α together with their uncertainties are listed for all investigated cases.

Finally, Fig. 8 shows the determined stochastic functions for all investigated cases. As in Fig. 4 the points with error bars stand for the numerically calculated values. The squares are averages of these analysis results. The solid curve represents the expected function according to system (40)–(42).

With the help of the numerically determined values for the deterministic and the stochastic parts of the dynamics a new time series $Y(t)$ has been integrated according to relation (38). The needed values for the deterministic and stochastic function have been found by interpolation or linear extrapolation. Figure 9 shows the original, analyzed time series $X(t)$ and the with the help of the analysis results reconstructed time series $Y(t)$ for the different Lévy indices α.

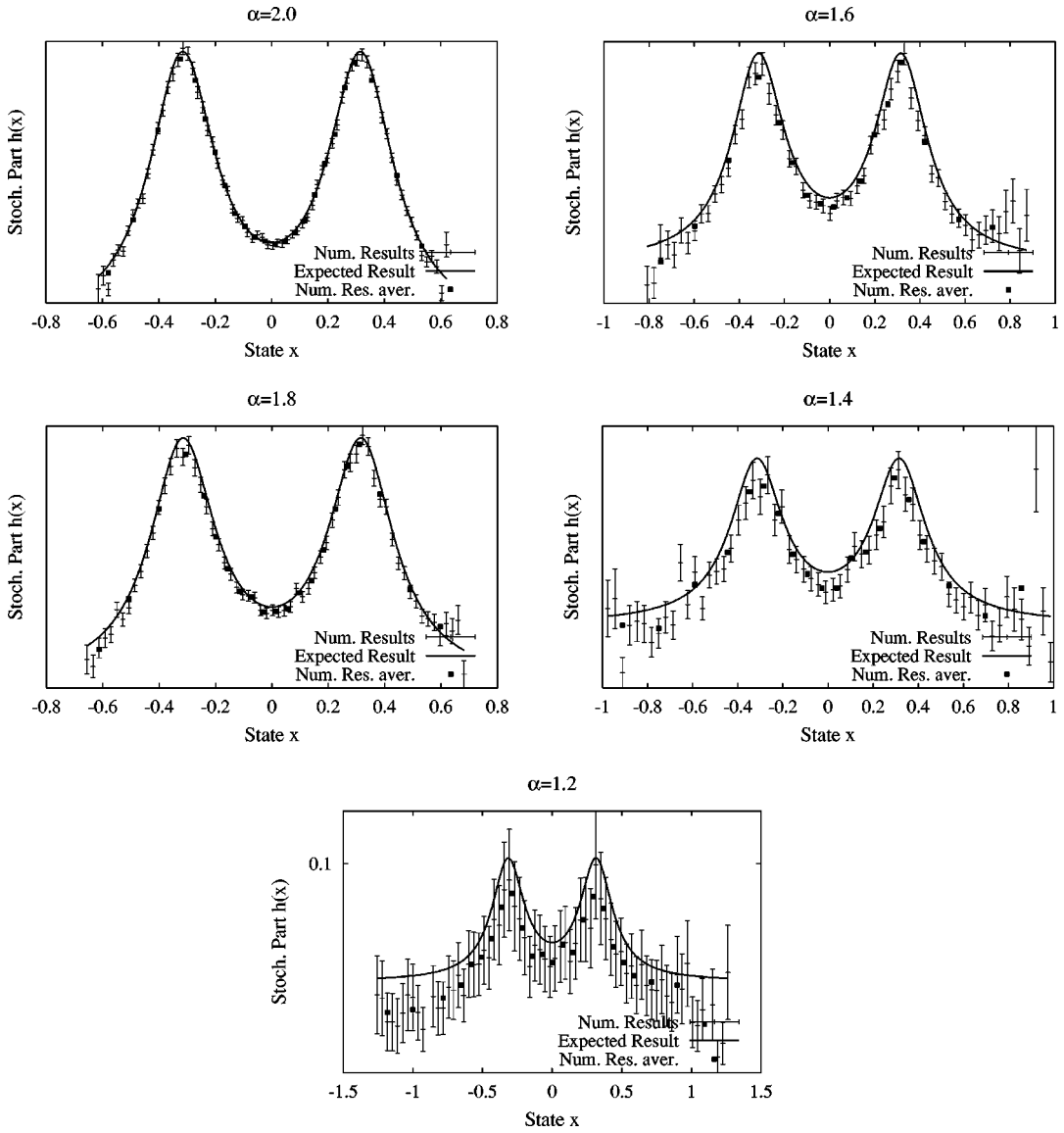

FIG. 8. Numerically determined values for the stochastic function $h(x)$ of system (40)–(42) in state space $\{x\}$ for different Lévy indices α. The values were calculated directly by data analysis of the time series shown in subsequences in Fig. 4, were smoothed and were compared with the theoretical curve.

As last step of the analysis, the assumption that the investigated system belongs to the class (16) of dynamical systems has to be validated. This is done by comparison of the conditional probability densities $p(x_1,t+\Delta t_{samp}|x_0,t)$ for different x_0 calculated on the one hand side by the measured trajectory and on the other hand by the reconstructed trajectory. For the Lévy index $\alpha=2.0$ (Gaussian process) Fig. 10 shows the probability distributions for some values of x_0. For calculating the distributions of the reconstructed time series 20 million points have been integrated.

VII. DISCUSSION OF THE RESULTS

In Figs. 5 and 8 numerical analysis results for the deterministic and stochastic parts of the investigated dynamical system are presented. Some interesting things concerning these results shall be remarked.

The x range in state space of the results increases with decreasing α. Numerical results can only be calculated for those x values in state space that are visited statistically often by the measured trajectory. Smaller Lévy indices α lead to larger, quite frequent deviations so that the area of the phase space that is covered by the investigated trajectory increases with decreasing Lévy index.

The greater the number of the trajectory's visits of a point x in state space is, the smaller is the uncertainty of the numerical results at this point. This explains why the error bars are bigger in the extreme regions of the covered and analyzed phase space than at the attracting points. The impression of smaller error bars in the deterministic parts for smaller Lévy indices is just caused by different scales of the coordinate axis.

If the behavior of the system shall be investigated in regions that are normally not visited by the trajectory, the sys-

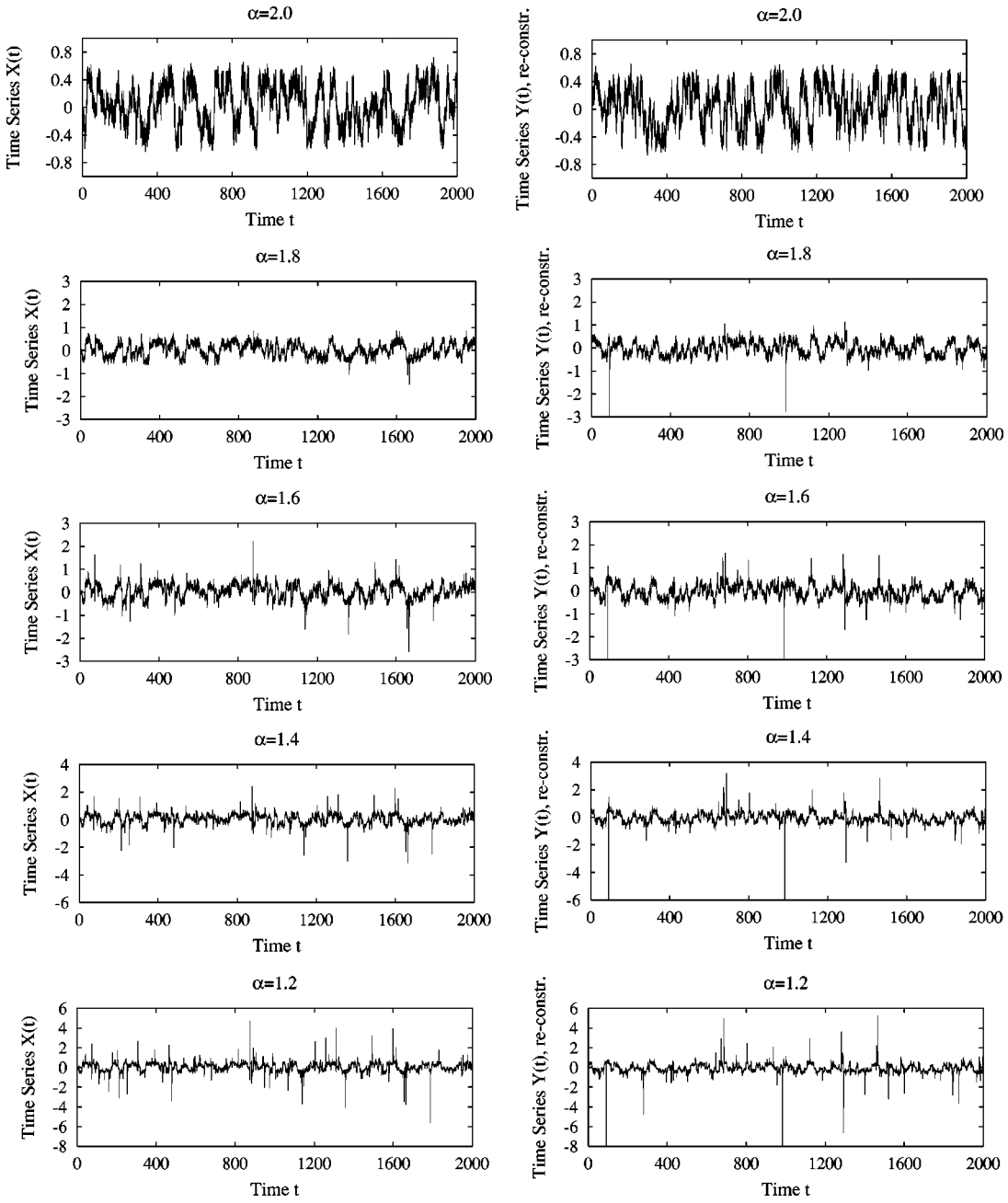

FIG. 9. Investigated time series $X(t)$, generated by iteration (42), over time t, in comparison with the reconstructed time series $Y(t)$ over time t, generated according to Eq. (38) with the numerically determined dynamical functions that have been found by data analysis.

tem has to be disturbed by additional dynamical noise in such a way that the trajectory covers the parts of state space of interest.

In Fig. 7 the determined values for α in the case $\alpha=1.6$ are plotted together with their uncertainties for different values of x. Again, it can be recognized that the uncertainty increases for extreme values of the covered state space because of a decreasing number of visits.

In Fig. 8 numerical results for the stochastic part of the investigated dynamical system are presented. Additionally to the comments above about both, the deterministic and stochastic results, one recognizes a strong increase of the uncertainties with smaller Lévy index α. This can be understood by relation (37).

In Fig. 9 subsequences of the analyzed time series and the reconstructed time series are plotted for different values of α. The reconstructed time series has been calculated with the help of the numerically determined values for the deterministic and the stochastic part and the Lévy index. But values for the state dependent functions exist only in those areas of the phase space that have been visited by the trajectory statistically often. In regions in the outer part results can only be found by extrapolation. According to the law of rare events this has been done in a linear way. When comparing the original and the reconstructed time series one recognizes a statistically identical behavior in the main part of the covered state space, but a slightly different behavior in strong deviations. The extreme values are taken by the recon-

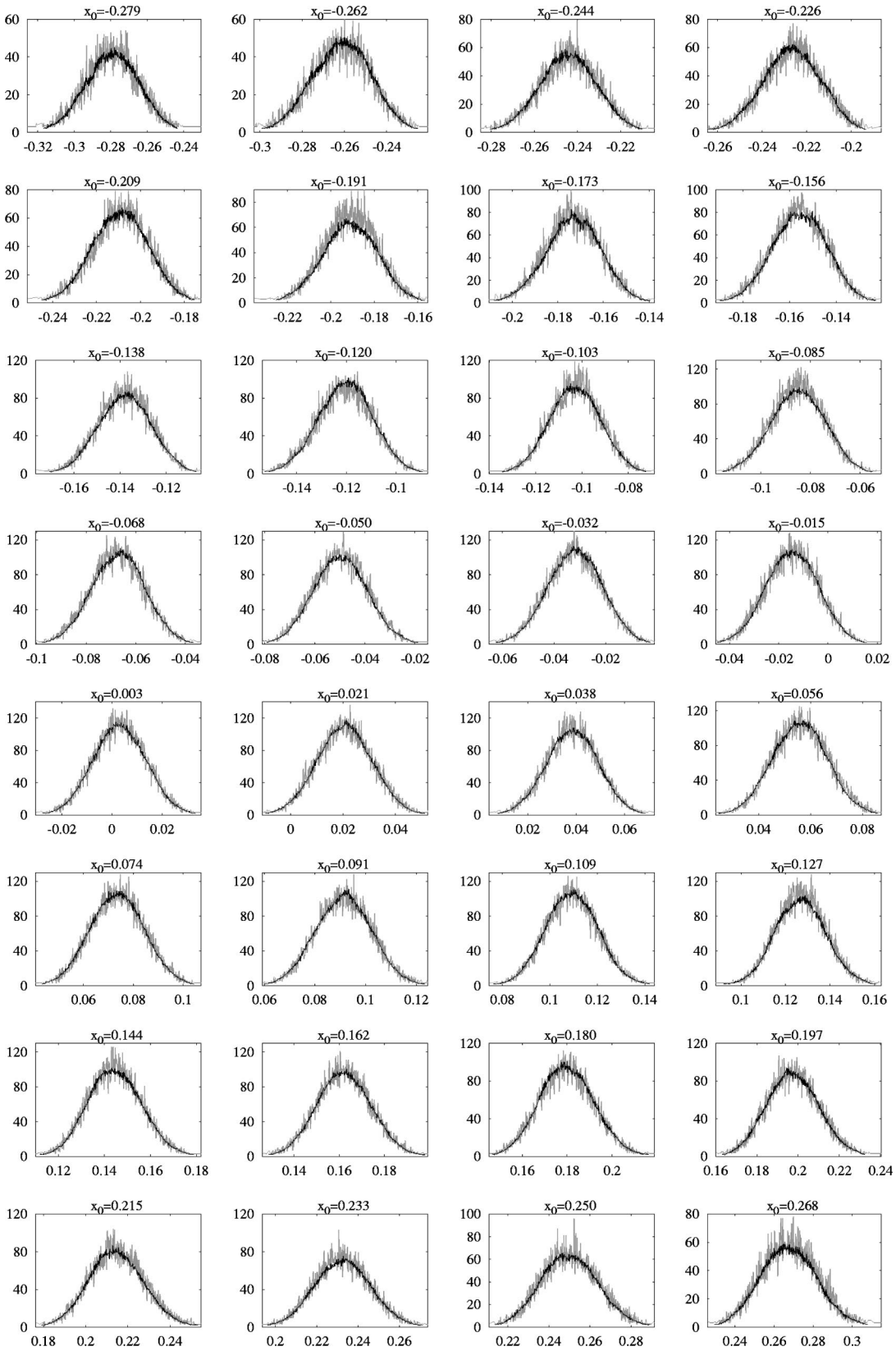

FIG. 10. Conditional probability density functions $p(x_1,t+\Delta t_{samp}|x_0,t)$ in arbitrary units over state x_1 for different values of x_0 for $\alpha=2.0$. The light curve belongs to the measured time series, the dark curve has been determined by the reconstructed time series.

structed trajectory statistically as often as by the original time series, but the recurrence to the main part of the state space is fulfilled in a slowlier way because of the linear extrapolation.

In Fig. 10 the functions of the probability density distribution $p(x_1,t+\Delta t_{samp}|x_0,t)$ are plotted for different values of x_0 for the Lévy index $\alpha=2.0$. A good agreement of the affiliated distributions can be recognized. The stronger fluctuations of the distributions of the measured trajectory for $\alpha=2.0$ in comparison with the distributions of the reconstructed trajectory are caused by the smaller number of data points (factor 20). The more data points are used for the calculation of the probability density distributions the smaller are the fluctuations.

VIII. SUMMARY

A method was presented that allows the data analysis of nonlinear Lévy systems. The Lévy index α can be determined as well as the deterministic and stochastic nonlinear parts of the dynamics. Model equations for the dynamical evolution of the investigated system can be formulated. Error bars show the uncertainties of the calculated values.

One remarkable feature of the algorithm is that no preknowledge about the system, no assumption or functional ansatz are necessary for the analysis. The only assumption, that the system belongs to the considered class of dynamical Lévy systems, is validated as last step of the procedure.

The fact that systems with nonlinear deterministic as well as nonlinear stochastic parts, can be investigated opens up the way to a lot of unknown, so far not investigated and modeled systems, physical systems as well as biological, medical or technical systems. Model equations that have been set up by logical reasons, symmetry or just experience can be validated or formulated in a more detailed way.

It was shown how time series can be reconstructed. By this way time series with any number of data points can be calculated. With the help of these reconstructed time series the long-time behavior of the investigated system can be simulated and characteristics whose calculation requires long series can be determined.

The presented method allows a data-driven formulation of model equations for the wide class of self-consistent nonlinear stochastic processes.

ACKNOWLEDGMENTS

The authors thank A. Zanker for fruitful discussions and the German foundation ''Studienstiftung des deutschen Volkes'' for their support of this work.

[1] C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd ed. (Springer-Verlag, Berlin, 1985).
[2] H. Risken, *The Fokker-Planck Equation* (Springer-Verlag, Berlin, 1989).
[3] J. Honerkamp, *Stochastische Dynamische Systeme* (VCH Verlagsgesellschaft, Weinheim, 1990).
[4] H. Haken, *Advanced Synergetics* (Springer-Verlag, Berlin, 1983).
[5] H. Haken, *Information and Self-Organization* (Springer-Verlag, Berlin, 1988).
[6] S. Siegert, R. Friedrich, and J. Peinke, Phys. Lett. A **243**, 275 (1998).
[7] R. Friedrich, S. Siegert, and J. Peinke, in *Transport and Structure*, edited by S. C. Müller, J. Parisi, and W. Zimmermann (Springer-Verlag, Berlin, 1999).
[8] R. Friedrich, S. Siegert, J. Peinke, St. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister, Phys. Lett. A **271**, 217 (2000).
[9] J. Gradisek, S. Siegert, R. Friedrich, and I. Grabec, Phys. Rev. E **62**, 3146 (2000).
[10] J. Gradisek, S. Siegert, R. Friedrich, and I. Grabec, in *Stochastic and Chaotic Dynamics in the Lakes*: Stochaos, edited by D. S. Broomhead, E. A. Luchinskaya, P. V. E. McClintock, and T. Mullin, AIP Conf. Proc. No. 502 (AIP, Melville, NY, 2000), pp. 476–481.
[11] P. Lévy, *Calcul des Probabilités* (Gauthier-Villars, Paris, 1925).
[12] P. Lévy, *Théorie de l'Addition des Variables Aléatoires* (Gauthier-Villars, Paris, 1937).
[13] W. Feller, *An Introduction to Probability Theory and Its Application* (Wiley, New York, 1957), Vol. 1; W. Feller, *An Introduction to Probability Theory and Its Application* (Wiley, New York, 1966), Vol. 2.
[14] H. C. Fogedby, Phys. Rev. Lett. **73**, 2517 (1994).
[15] S. Jespersen, R. Metzler, and H. C. Fogedby, Phys. Rev. E **59**, 2736 (1999).
[16] G. Zumofen and J. Klafter, Phys. Rev. E **51**, 2805 (1995).
[17] O. V. Bychuk and B. O'Shaughnessy, Phys. Rev. Lett. **74**, 1795 (1995).
[18] S. Stapf, R. Kimmich, and R.-O. Seitter, Phys. Rev. Lett. **75**, 2855 (1995).
[19] J. Bodurka, R.-O. Seitter, R. Kimmich, and A. Gutsze, J. Chem. Phys. **107**, 5621 (1997).
[20] A. Ott, J. P. Bouchaud, D. Langevin, and W. Urbach, Phys. Rev. Lett. **65**, 2201 (1990).
[21] G. Zumofen, A. Blumen, J. Klafter, and M. F. Shlesinger, J. Stat. Phys. **54**, 1519 (1989).
[22] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. Murphy, P. A. Prince, and H. E. Stanley, Nature (London) **381**, 413 (1996).
[23] J. Klafter, A. Blumen, and M. F. Shlesinger, Phys. Rev. A **35**, 3081 (1987).
[24] H. C. Fogedby, Phys. Rev. E **50**, 1657 (1994).
[25] R. Metzler, E. Barkai, and J. Klafter, Europhys. Lett. **46**, 431 (1999).
[26] G. Samorodnitzky and M. Taqqu, *Stable Non-Gaussian Random Processes* (Chapman & Hall, New York, 1994).
[27] J. Honerkamp, *Statistical Physics* (Springer-Verlag, Berlin, 1998).