

Quantifying model uncertainty for the observed non-Gaussian data by the Hellinger distance

Yayun Zheng¹, Fang Yang^{*2}, Jinqiao Duan³, Jürgen Kurths^{4,5}

¹ School of Mathematical Sciences, Jiangsu University, Zhenjiang 212013, China.

² School of Mathematics and Statistics and Center for Mathematical Science, Huazhong University of Science and Technology, Wuhan, 430074, China.

³ Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA.

⁴ Potsdam Institute for Climate Impact Research, Potsdam, 14473, Germany.

⁵ Department of Physics, Humboldt University, Berlin, 12489, Germany.

⁶ Centre for Analysis of Complex Systems, Sechenov First Moscow State Medical University, Moscow, Russia.

Abstract

Mathematical models for complex systems under random fluctuations often contain uncertain parameters. However, quantifying model uncertainty for a stochastic differential equation with an α -stable Lévy process is still lacking. Here, we propose an approach to infer all the uncertain non-Gaussian parameters and other system parameters by minimizing the Hellinger distance over the parameter space. The Hellinger distance measures the similarity between an empirical probability density of non-Gaussian observations and a solution (as a probability density) of the associated nonlocal Fokker-Planck equation. Numerical experiments verify that our method is feasible for estimating single and multiple parameters. Meanwhile, we find an optimal estimation interval of the estimated parameters. This method is beneficial for extracting governing dynamical system models under non-Gaussian fluctuations, as in the study of abrupt climate changes in the Dansgaard-Oeschger events.

Keywords: Non-Gaussian observations; Parameters estimation; Hellinger distance; Probability density

1 Introduction

Complex systems under influences of random fluctuations also have uncertain parameters [1]. An important problem in modeling such random processes by stochastic differential equations (SDEs) is to estimate uncertain parameters from observations of the stochastic paths.

^{1*} Corresponding author.

E-mail address: yfang@hust.edu.cn (Fang Yang)

A Brownian motion has properties of continuous sample paths, normal diffusion and light tail (probability density decays exponentially), theoretical results on parametric estimations for SDEs driven by Brownian motion are relatively well developed. The Gaussian kernel density estimator [2, 3] and the Bayesian estimator [4, 5] are well-known approaches for parameters estimation of a drift function when the observations of complete paths are available. Furthermore, by the nonparametric estimation method of Kramers-Moyal coefficients [6], the statistical definitions of conditional first and second moments [7] or variational formulation of the stationary Fokker-Planck equation [8], we could provide an expression for the drift function and the diffusion one.

However, various complex phenomena involve non-Gaussian fluctuations, with properties such as intermittent jumps, anomalous diffusion, and heavy tail (probability density decays with power law) distribution. A heavy-tailed distribution, like a Lévy distribution, is characterized by a high likelihood for extreme events, compared to a normal distribution. For instance, Ditlevsen shows that the paleoclimatic records for Dansgaard-Oeschger events have a strong non-Gaussian distribution [9]. The protein production occurs in bursts which are observed during a genetic regulation [10]. Meanwhile, experimental studies find that Lévy flights are an optimal pattern when the prey is sparsely and randomly distributed for open-ocean predatory fish [11]. The Lévy process is also used in other scientific domains, for example, it has been shown that certain stock price has statistical properties that are compatible with a Lévy random walk [12]. Additionally, in the field of cognitive research, a few studies provide evidence of Lévy processes, e.g., to search and cluster in semantic memory [13] and human decision making [14]. An α -stable Lévy process is thought to be an appropriate model for a non-Gaussian heavy-tailed process. Therefore, for modeling these complex systems, it becomes necessary and significant to consider parameters estimation for a stochastic system driven by the α -stable Lévy process. In general, the p -th moment of an α -stable Lévy random variable is finite if and only if $p < \alpha$ ($0 < \alpha < 2$), so it does not have second moments. Meanwhile, the stationary probability density of an α -stable Lévy process does not always exist. Due to these disadvantages, unfortunately, the existing methods for parameter estimation of Brownian motion are not applied for dynamical system with non-Gaussian fluctuations.

There are few results about the parametric estimation for stochastic processes driven by Lévy processes. In some special cases, it is possible to infer parameters only for the drift function assuming that the values of other parameters are known. For example, a simple Ornstein-Uhlenbeck process is considered, i.e., the drift function is known to be linear, or stochastic processes are driven by a compound Poisson process [15]. In these works, the quasi-maximum likelihood, a self-weighted least absolute deviation estimator [16] or trajectory fitting estimator are established for discretely observed Lévy processes. For an α -stable Lévy process, the parametric estimation problem becomes more difficult because the second moment does not exist. Recently, Hu and Long *et al.* [17, 18] addresses a trajectory fitting and a least-square estimator on estimation of a drift parameter for a stochastic system under an α -stable Lévy noise. Fasen [19] extended the results to high dimensions.

In the above-mentioned works, one can only estimate the drift parameters. Meanwhile, the non-Gaussian index α plays a decisive role in the construction of Lévy processes. An α -stable Lévy process has larger jumps with lower jump probabilities when α is small ($0 < \alpha < 1$), while it has smaller jumps with higher jump frequencies for large α values ($1 < \alpha < 2$). The special cases for $\alpha = 1$ and $\alpha = 2$ correspond to the Cauchy process and the Brownian motion, respectively. Therefore, the estimation of the parameter

α is extremely important. There are some simple and straightforward approaches to learn this α from the path observation, such as the slope of the log-log linear regression [20] or the Hill estimator [21]. These methods do not assume a parametric form for the entire distribution function, but focus only on the tail behavior. However, the true tail behavior of Lévy distribution is visible only for extremely large data sets, or it is a challenge to choose the right value of the largest order statistics.

There have been no available estimators simultaneously for the drift parameter and other α -stable Lévy parameters, including α and other non-Gaussian parameters. An alternative method is relied on the characteristic function of α -stable Lévy process [22]. Based on the ergodic theory and sample characteristic functions, Cheng *et al.* [23] study a Ornstein-Uhlenbeck process with the α -stable Lévy noise. The parameter estimation for α and the other parameters is obtained by matching the empirical characteristic function with the corresponding theoretical one. We note that a method of numerical optimization is devised in [24], where two deterministic quantities: mean exit time or the escape probability is observed to estimate the uncertain parameter and other system parameters. It is based on solving an inverse problem for a deterministic, nonlocal partial differential equation.

The existing works provide certain approaches to estimate the drift parameter and other α -stable Lévy parameters. With severe limitations, (i) the drift term can only be a linear function; (ii) the empirical characteristic function is approximately defined; (iii) it is difficulty to observe mean exit time or first escape probability from the discrete time series data. In response to the existing challenge, we are interested in finding an effective and feasible approach for parameter estimation of a stochastic system under an α -stable Lévy noise. The method can be applied for a nonlinear drift term, α and other system parameters can be estimated simultaneously. Compared with the characteristic function and other quantities, a probability density or probability distribution becomes easy to obtain from an observation data set with heavy-tailed distribution.

We recall some recent works on estimating parameters of α -stable Lévy processes based on probability densities. Chen and Chen [25] choose a mixture of Cauchy and Gaussian distribution to approximate the probability density function of the α -stable Ornstein-Uhlenbeck distribution. By means of transition function and Laplace transform, they construct an explicit approximate sequence of the maximum likelihood function to obtain the estimation of parameters. Inspired by the derivation of the differential Chapman-Kolmogorov equation, Li and Duan [26] derive Kramer-Moyal formulas to express the jump measure, drift and diffusion coefficient of a stochastic differential equation with respect to the transition probability density .

The probability density function for SDE driven by an α -stable Lévy process satisfies a deterministic, nonlocal differential equation with an initial condition, i.e., nonlocal Fokker-Planck equation, which have a nonlocal or fractional Laplacian term. In terms of theory, we derived the Fokker-Planck equations for Marcus SDEs driven by Lévy processes in high dimensional [27]. In terms of numerical calculations, taking advantage of the Toeplitz matrix structure of the time-space discretization, Gao *et al.* [28] proposed a fast and accurate numerical algorithm to simulate nonlocal Fokker-Planck equations under either absorbing or natural conditions. Meanwhile, a piecewise integro quadratic spline interpolation approach [29-31] and a finite element method [32] are developed for the approximate nonlocal or fractional integral.

Consequently, we propose an approach to infer simultaneously the drift parameter and other α -stable Lévy parameters. Our method is based on minimizing the Hellinger distance between the observed proba-

bility distribution and the solution (as a probability distribution) of the associated Fokker-Planck equation for a general stochastic dynamical system driven by an α -stable Lévy process.

In the present paper, we consider the parameter estimation problem of an α -stable Lévy stochastic dynamical system containing uncertain parameters. In Section 2, we propose a method of estimating the uncertain parameters based on the Hellinger distance of the probability densities. In Section 3, we present some simulation results of estimation for single and multiple parameters by minimizing the Hellinger distance. Finally we give some concluding and future works in Section 4.

2 Methods

We consider a dynamical system with heavy-tailed uncertainty, which could be modeled by a stochastic process $X(t)$

$$dX(t) = f(X(t), \theta)dt + \epsilon dL^\alpha(t), \quad X(0) = x_0 \in \mathbb{R}^1, \quad (1)$$

where the drift function $f(x, \theta)$ has the uncertain parameter θ , and a scalar symmetric α -stable Lévy process $L^\alpha(t)$ with the non-Gaussian index $0 < \alpha < 2$ is defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The parameter ϵ is the non-negative α -stable Lévy noise intensity.

A scalar symmetric α -stable Lévy process is characterized by a generating triplet (b, Q, ν_α) , a linear coefficient b , a diffusion parameter Q , and a nonnegative Borel measure ν_α . This jump measure ν_α is defined on $(\mathbb{R}^1, \mathfrak{B}(\mathbb{R}^1))$ [33] by:

$$\nu_\alpha = \frac{C_\alpha dy}{|y|^{1+\alpha}},$$

with $0 < \alpha < 2$ and $C_\alpha = \frac{\alpha}{2^{1-\alpha}\sqrt{\pi}} \frac{\Gamma(\frac{1+\alpha}{2})}{\Gamma(1-\frac{\alpha}{2})}$. In this paper, we consider an α -stable Lévy process with a triplet $(0, 0, \nu)$, i.e., a pure jump process.

For $0 < \alpha < 2$, the α -stable Lévy process $L^\alpha(t)$ has a heavy-tailed distribution [34]

$$\mathbb{P}(|L^\alpha(t)| > y) \sim \frac{1}{y^\alpha},$$

as the tail estimate decays in a power law. Therefore α is also called the power parameter. The tail behavior is different from the Brownian motion with light tail, as the tail decays exponentially.

We assume that the drift term f is local Lipschitz continuous. Then the SDE (1) has a unique solution [33]. The conditional probability density $p(x, t|x_0, 0) \triangleq p(X(t) = x|X(0) = x_0)$ represents the density of the $X(t)$ given a value x_0 at initial time. For convenience, we drop the initial condition and simply denote it by $p(x, t)$. There exists sufficient condition for the existence and regularity of the probability density $p(x, t)$ for some SDEs driven by Lévy processes. The existence is based on Malliavin calculus with jumps under Hörmander's condition, see Refs. [35-37] and the references therein for more details.

We see that the stochastic process X_t in Eq.(1) under Lévy noise depends on the following parameters. The first one is an uncertain system parameter θ . In general, the estimated parameter θ plays a key

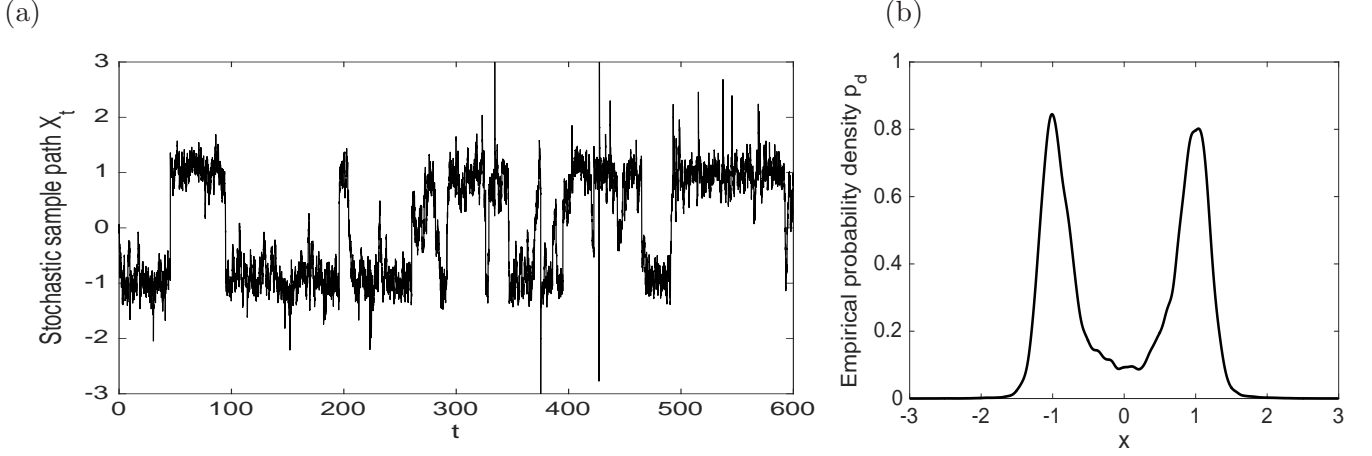


Figure 1: (a) The stochastic sample path of the process X_t (Eq.(4)) with $\alpha = 1.7$, $\epsilon = 0.3$ and $\theta = 1$. (b) The empirical probability density p_d of the stochastic trajectory (a).

role in the system model, which could be a bifurcation parameter inducing a transition between states. The control parameter θ could be a greenhouse factor in the case of the energy balance model [38], or a freshwater forcing strength in the thermohaline circulation one [39]. Besides, there are uncertain Lévy parameters: the non-Gaussian index α and the Lévy noise intensity ϵ .

Let us assume that we have access to a set of observations $y = (y_1, y_2, \dots, y_n)$, which are the version of the process X_t with a non-Gaussian distribution sampled at discrete times $t_k \in [0, T]$ for $k = 1, 2, \dots, n$, i.e., $y_k = X_{t_k}$ for $k = 1, 2, \dots, n$. In this paper, we will discuss the problem of estimating the parameters α , ϵ and θ simultaneously using the observations $y = (y_1, y_2, \dots, y_n)$.

To reach this purpose, we would like to introduce the Hellinger distance. It is used to quantify the similarity between two probability distributions [40]. The Hellinger distance between two probability density functions $p(x)$ and $p_d(x)$ is

$$H^2(p, p_d) = \frac{1}{2} \int_{\mathbb{R}^1} \left(\sqrt{p(x)} - \sqrt{p_d(x)} \right)^2 dx.$$

The Hellinger distance H satisfies the property: $0 \leq H(p, p_d) \leq 1$. Here, p_d is the empirical probability density from an observation data set $y = (y_1, y_2, \dots, y_n)$. The probability density function $p(x)$ is a solution of the nonlocal Fokker-Planck equation $p(x, t)$ at time t .

$$\frac{\partial}{\partial t} p(x, t) = -\frac{\partial}{\partial x} (f(x, \theta) p(x, t)) + \epsilon^\alpha \int_{\mathbb{R}^1 \setminus \{0\}} \left[p(x + y, t) - p(x, t) - I_{|y| < 1}(y) y \frac{\partial}{\partial x} p(x, t) \right] \nu_\alpha(dy). \quad (2)$$

The integral part in the right hand side is actually the nonlocal Laplacian operator. This nonlocality is the manifestation of effect of non-Gaussian Lévy fluctuations [41]. The equation fulfills an initial condition

$$\lim_{t \rightarrow 0} p(x, t | x_0, 0) = \delta(x - x_0).$$

We consider that the observation set comes from an α -stable Lévy distribution $p(x, t)$. Associated with each probability density is the parameters set $\lambda = (\theta, \alpha, \epsilon) \in \Theta$, where Θ is called the parameter space, a finite-dimensional subset of the Euclidean space. Evaluating the Hellinger distance at the observed data set y gives an objective function

$$G(\lambda) = H^2(p(x, \lambda), p_d(x)).$$

The Hellinger distance estimation aims to find the value of the model parameters that minimize the objective function over the parameter space Θ , that is

$$\hat{\lambda} = \arg \min_{\lambda \in \Theta} G(\lambda).$$

To address the probability density $p(x, \lambda)$, we use the numerical algorithm of Gao *et al.* [28] to solve the nonlocal differential equation in Eq.(2) under the absorbing condition. This absorbing condition means that the probability of finding “partical” X_t outside the finite interval $D = (a, b)$ is zero. We decompose the integral part of Eq.(2) into three parts $\int_{\mathbb{R}^1} = \int_{-\infty}^{a-x} + \int_{b-x}^{a-x} + \int_{b-x}^{\infty}$ in \mathbb{R}^1 and analytically evaluate the first and third integrals, then Eq.(2) changes to

$$\begin{aligned} \frac{\partial}{\partial t} p(x, t) = & -\frac{\partial}{\partial x} (f(x)p(x, t)) - \frac{\epsilon^\alpha C_\alpha}{\alpha} \left[\frac{1}{(x-a)^\alpha} + \frac{1}{(b-x)^\alpha} \right] p(x, t) \\ & + \epsilon^\alpha C_\alpha \int_{a-x}^{b-x} \frac{p(x+y, t) - p(x, t) - I_{|y|<1}(y)y \frac{\partial}{\partial x} p(x, t)}{|y|^{1+\alpha}} dy, \end{aligned} \quad (3)$$

for $x \in (a, b)$. The non-Gaussian index $\alpha \in (0, 2)$ and the Lévy intensity $\epsilon \in (0, 1]$.

3 Numerical Experiments

We now explore how the Hellinger distance can be used to estimate the parameters of SDE driven by a symmetric α -stable Lévy process. We consider the following example

$$dX(t) = (-\theta X(t)^3 + X(t))dt + \epsilon dL^\alpha(t), \quad X(0) = 0 \in \mathbb{R}^1, \quad (4)$$

In this example, a nonlinear drift term is $f(x, \theta) = -\theta x^3 + x$ with uncertain parameter θ . We start with training data from numerical simulations of Eq.(4). The stochastic trajectory can be regarded as a heavy-tailed time series with the parameters $\alpha = 1.7$, $\theta = 1$ and $\epsilon = 0.3$ shown in the Fig. 1(a). Here, the choice of parameters $\alpha \in (0, 2)$, $\theta \in [0.5, 1.5]$ and $\epsilon \in (0, 1]$ are arbitrary. Our work focus on the comparison the similarity between the estimated and original parameters in the SDE (4) by minimizing the Hellinger distance. The Hellinger distance measures the similarity between an empirical probability density of non-Gaussian observations and a solution (as a probability density) of the associated nonlocal Fokker-Planck equation (2).

The empirical probability density function p_d for observations (Fig. 1(a)) could be determined by the normal kernel method. In the simulation, we use the MATLAB function *ksensity* to evaluate the $p_d(x)$

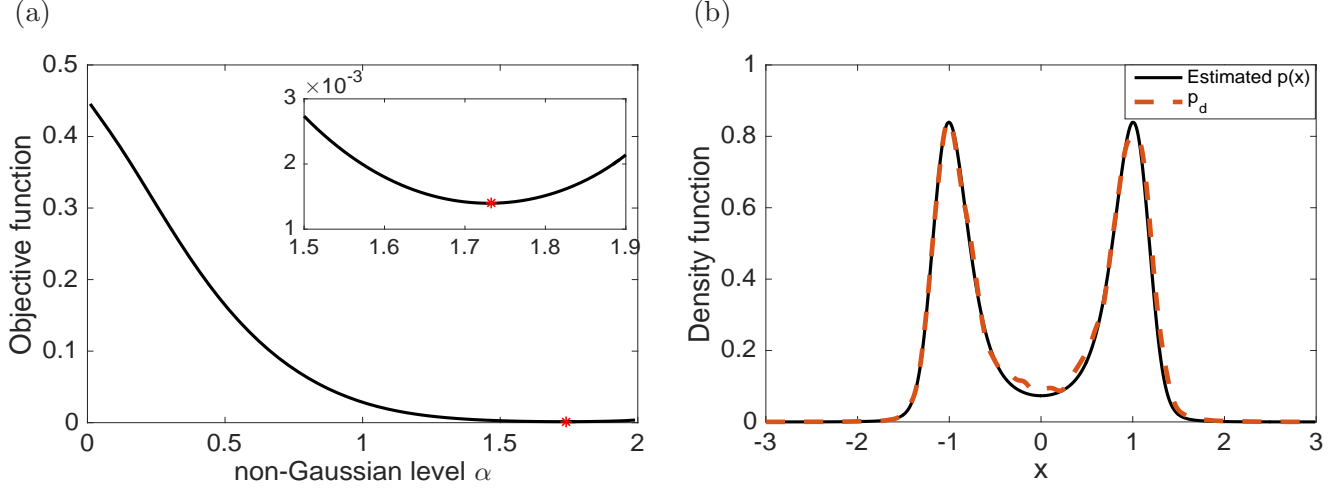


Figure 2: (a) The Hellinger distance estimation of α . The estimated value $\hat{\alpha} = 1.738$ corresponds to the minimum objective function $G_{\alpha} = 0.0014$ (red \star). (b) The estimated density function $p(x)$ with $\hat{\alpha} = 1.738$, $\theta = 1$ and $\epsilon = 0.3$ compared with the empirical density p_d in Fig.1(b).

for $x \in (-3, 3)$ as shown in Fig. 1(b). A selected bandwidth is $h = 1.8s/n^{1/5}$, where $n = 6 \times 10^6$ is the number of observed data points and s is the standard deviation of the data set.

Now we provide the details of the computation to infer parameters from the observations. The value of the parameters are estimated by minimizing the Hellinger distance over the parameter space Θ . We shall first estimate single parameter assuming that the values of the other parameters are known, and then estimate multiple parameters

3.1 Estimation for a single parameter

We want to find out an estimation of α by achieving a numerical optimization of the objective function G of the Hellinger distance. We consider the parameter $\lambda = \alpha$ is in the parameter space $\Theta = (0, 2) \subset \mathbb{R}^1$, and assume that the other parameters are known, i.e. $\epsilon = 0.3$ and $\theta = 1$. Then the objective function G of the Hellinger distance is

$$G(\alpha) = \frac{1}{2} \int_{\mathbb{R}^1} \left(\sqrt{p(x, \alpha)} - \sqrt{p_d(x)} \right)^2 dx.$$

Based on the numerical algorithm of Gao *et al.* [28], the probability density $p(x, \alpha)$ is solved by the nonlocal differential equation (2) given $\alpha \in (0, 2)$ for $x \in D = (-3, 3)$ at $t = 50$. In the numerical simulations, the probability profile of its initial position is Gaussian $p(x, 0) = \sqrt{\frac{40}{\pi}} e^{-40x^2}$. We have chosen the spatial resolution $h = 0.003$ and the time step size $\Delta t = 0.5h^2$.

In Fig. 2(a), we employ a discretization step of $\Delta\alpha = 0.035$ and use 55 grid points for $\alpha \in (0, 2)$. Then the estimation of $\hat{\alpha} = 1.738$ is obtained with the minimum value of the Hellinger distance $G(\hat{\alpha}) = 0.001399$

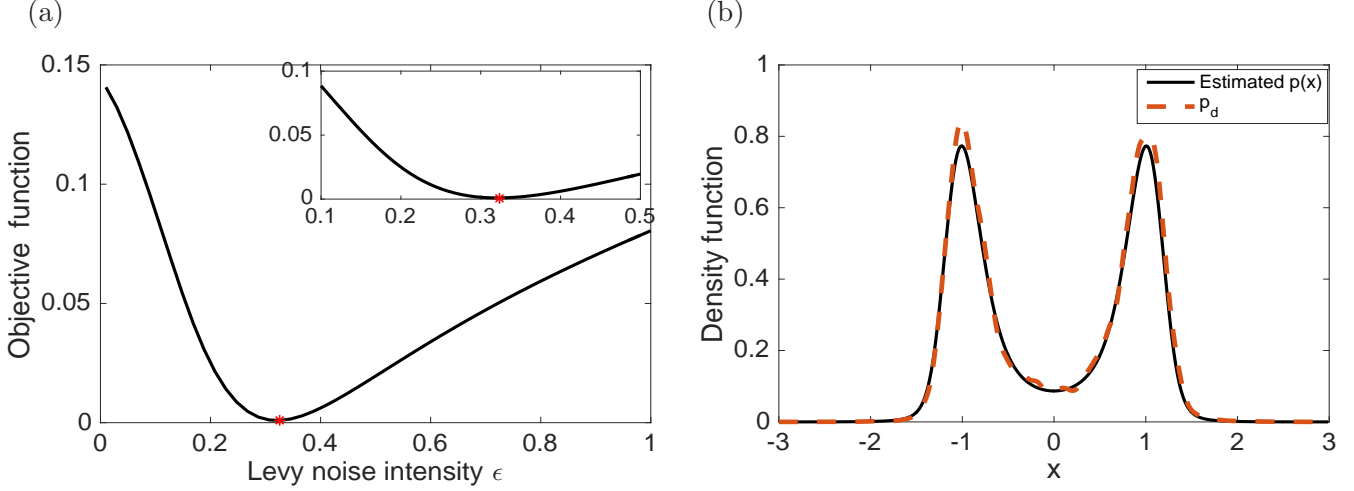


Figure 3: (a) The Hellinger distance estimation of ϵ . The estimated value $\hat{\epsilon} = 0.3268$ corresponds to the minimum objective function $G(\hat{\epsilon}) = 0.000951$ (red \star). (b) The estimated density function $p(x)$ with $\hat{\epsilon} = 0.3268$, $\theta = 1$ and $\alpha = 1.7$ compared with the empirical density p_d in Fig.1(b).

over the parameter space $\Theta = (0, 2)$. Furthermore, we restrict α on a small region $[1.5, 1.9]$ for accurately estimation. The result illustrates that the estimated values of $\hat{\alpha} \in [1.684, 1.764]$ contains the true value of $\alpha = 1.7$ with the Hellinger distance $G(\hat{\alpha}) < 0.00145$ (inset figure in Fig. 2(a)). It means that we could find an optimal interval for the estimated parameter α . As an illustration, we show the results of the probability density $p(x, \alpha)$ of SDE (4) with estimated value $\hat{\alpha} = 1.738$ (dashed) and the empirical density p_d from the observed data (dotted) in Fig. 1(b). We can see that the estimated probability density presents the goodness-of-fit to the empirical one.

Similarity, we explore the dependence of the objective function G on the value of ϵ keeping the other parameters fixed. Fig. 3(a) shows the minimized the $G(\hat{\epsilon}) = 0.0009$ of the Hellinger distance that corresponds to the estimated value of the Lévy noise intensity $\hat{\epsilon} = 0.3268$. Meanwhile, we could get the optimal estimation interval $\hat{\epsilon} \in [0.292, 0.348]$ for the Hellinger distance $G(\hat{\epsilon}) < 0.0019$. This domain includes the true value of $\epsilon = 0.3$. In Figure 3(b), the empirical density p_d is well fitted by the probability density $p(x, \epsilon)$ with $\hat{\epsilon} = 0.3268$.

We have inferred the parameters α and ϵ by considering the Hellinger distance, respectively. Next, we would like to compare the Hellinger distance with other commonly used metrics, such as the L^2 norm distance, the maximum absolute error distance and the Sørensen distance to quantify the similarity between two probability distributions. The objective function of the L^2 norm distance is defined as

$$G(\lambda) = \frac{\|p(\lambda, x) - p_d(x)\|_2^2}{\|p_d(x)\|_2^2}.$$

While the maximum absolute approximation distance is

$$G(\lambda) = \max |p(\lambda, x) - p_d(x)|.$$

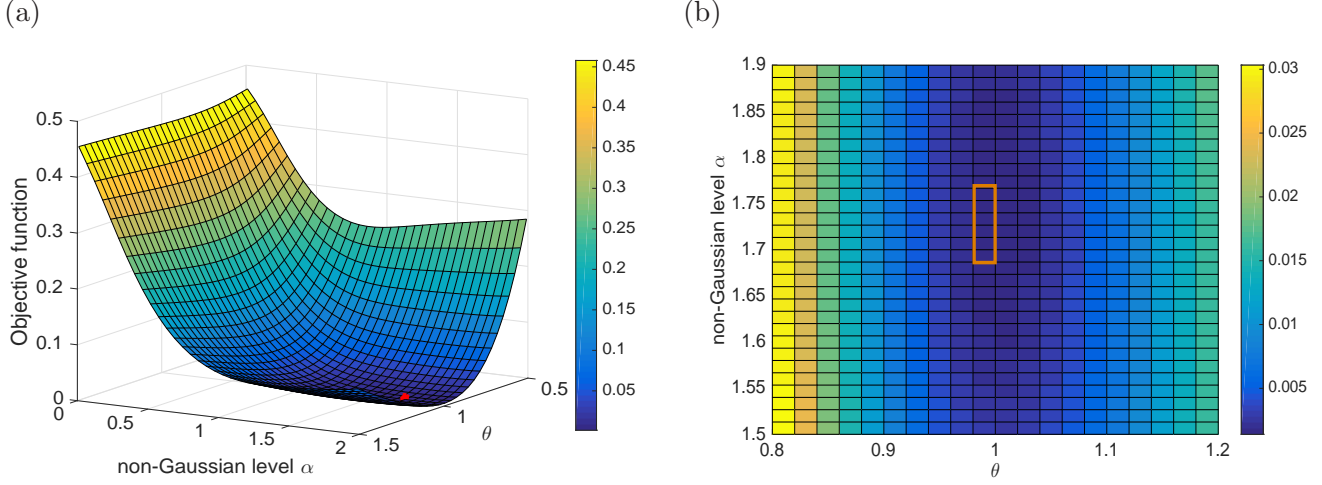


Figure 4: (a) Estimation of α and θ by optimization of the Hellinger distance. The estimated values are $\hat{\alpha} = 1.707$ and $\hat{\theta} = 0.9828$ with the minimum $G(\hat{\alpha}, \hat{\theta}) = 0.0014$ (red \star). (b) The optimal estimation domain of $\hat{\alpha}$ and $\hat{\theta}$ is $[1.6867, 1.78] \times [0.98, 1]$.

The Sørensen distance is used in ecology model. The expression of the objective function for the Sørensen distance is

$$G(\lambda) = \frac{\|p(\lambda, x) - p_d(x)\|_1}{\|p(\lambda, x) + p_d(x)\|_1}.$$

Here the $\|z\|_n = (|z_1|^n + |z_2|^n + \dots + |z_m|^n)^{1/n}$ is defined as n -norm of $z = (z_1, z_2, \dots, z_m)$. The estimation of uncertain parameters set λ could be achieved by minimizing G , i.e., $\hat{\lambda} = \arg \min G(\lambda)$ for $\lambda \in \Theta$.

Next, let us examine the effect of these kinds of distances on the estimation of α and ϵ , respectively. We keep the other parameters and the divided subintervals the same as those in Figs. 2 and 3. In Table 1, the results on these distance show that the Hellinger distance gives a better estimation for α than the others. In contrast, all four distances show a good fit to the true ϵ . In this example, the Hellinger distance is the most effective method to estimate parameters.

Distance	True value α	Estimated $\hat{\alpha}$	True value ϵ	Estimated $\hat{\epsilon}$	G_α	G_ϵ
Hellinger	1.7	1.7380	0.3	0.3268	0.0014	0.0009
L^2 norm	1.7	1.8100	0.3	0.3070	0.0028	0.0032
Maximum absolute error	1.7	1.9180	0.3	0.3070	0.0427	0.062
Sørensen	1.7	1.7740	0.3	0.3070	0.0299	0.0295

Table 1: Compared with different distances: Hellinger distance, L^2 norm distance, maximum absolute error distance and the Sørensen distance.

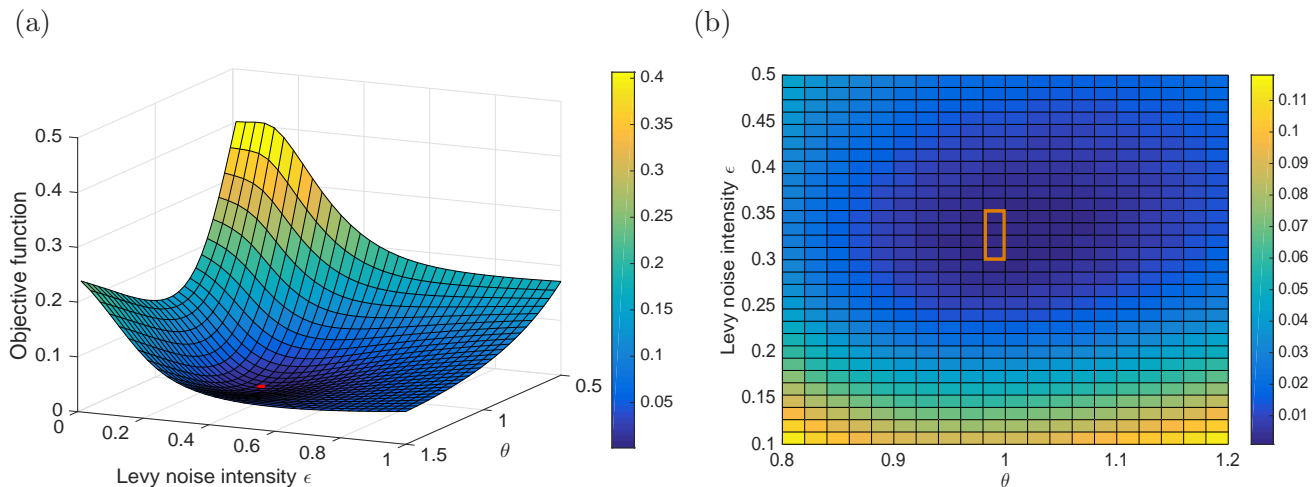


Figure 5: (a) Estimation of ϵ and θ . The estimated values are $\hat{\epsilon} = 0.3303$ and $\hat{\theta} = 1.0185$ with the minimum $G(\hat{\epsilon}, \hat{\theta}) = 0.0012$ (red \star). (b) The optimal estimation domain is $[0.3, 0.3267] \times [0.9, 1.02]$.

3.2 Estimation for multiple parameters

The above example has verified that our method is feasible for estimating a single parameter by minimizing the Hellinger distance. Next, we will apply this approach to estimate the multiple unknown parameters. First, we simplify our model by assuming that one parameter $\epsilon = 0.3$ is known and then estimate α and θ , while keeping the other factors the same as in the section 3.1. The objective function G is given by Hellinger distance

$$G(\alpha, \theta) = \frac{1}{2} \int_{\mathbb{R}^1} \left(\sqrt{p(x, \alpha, \theta)} - \sqrt{p_d(x)} \right)^2 dx.$$

The probability density $p(x, \alpha, \theta)$ is a solution of the nonlocal differential equation (2) given values of $\alpha \in (0, 2)$ and $\theta \in [0.5, 1.5]$ at $t = 50$. Fig. 4(a) shows that the objective function G changes with the values of α and θ in the parameter space $\Theta = (0, 2) \times [0.5, 1.5] \subset \mathbb{R}^2$. The minimum value of $G(\hat{\alpha}, \hat{\theta}) = 0.0014$ is identified with $\hat{\alpha} = 1.707$ and $\hat{\theta} = 0.9828$. In the same manner, the optimal estimation domain of $\hat{\alpha}$ and $\hat{\theta}$ is $[1.6867, 1.78] \times [0.98, 1]$ (orange rectangular frame) by further restricting the range of parameters as shown in Fig. 4(b).

Parameter	True value	Estimated $\hat{\lambda}_1 = (\hat{\epsilon}, \hat{\theta})$	Estimated $\hat{\lambda}_2 = (\hat{\alpha}, \hat{\epsilon})$
θ	1.0	1.0185	—
α	1.7	—	1.707
ϵ	0.3	0.3303	0.307
G	0	0.0012	0.0011

Table 2: Estimation of $\lambda_1 = (\epsilon, \theta)$ and $\lambda_2 = (\alpha, \epsilon)$.

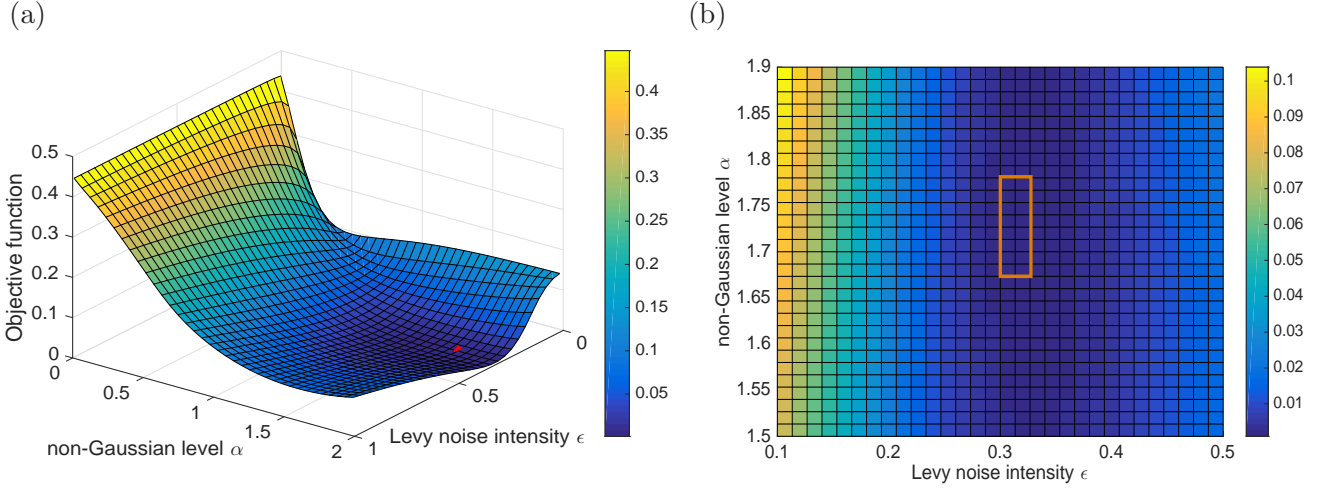


Figure 6: (a) Estimation of α and ϵ . The estimated values are $\hat{\alpha} = 1.707$ and $\hat{\epsilon} = 0.307$ with the minimum $G(\hat{\alpha}, \hat{\epsilon}) = 0.0011$ (red \star). (b) The optimal estimation domain is $[1.6733, 1.78] \times [0.3133, 0.3267]$.

Second, we take into account the estimation of the other two combinations of all three parameters, $\lambda_1 = (\epsilon, \theta)$ and $\lambda_2 = (\alpha, \epsilon)$ corresponding to the parameter spaces $\Theta_1 = (0, 1] \times [0.5, 1.5]$, $\Theta_2 = (0, 2) \times (0, 1]$, respectively. The estimated results are found by the minimized the Hellinger distance as shown in Table 2. Meanwhile, we can also determine the optimal domains of the estimated parameters sets $\hat{\lambda}_1$ and $\hat{\lambda}_2$ as shown in Figs. 5(b) and 6(b). The result shows that the Hellinger distance $G(\hat{\alpha}, \hat{\epsilon}) \leq 0.001$ if the estimated parameters set $\hat{\lambda}_1$ belongs to the domain $[1.6733, 1.78] \times [0.3, 0.3267]$. Meanwhile, the estimation domain of $\hat{\lambda}_2$ is $[0.3, 0.34] \times [0.98, 1]$ if the Hellinger distance $G(\hat{\epsilon}, \hat{\theta}) \leq 0.0015$.

Finally, we seek all these parameters $\lambda = (\alpha, \epsilon, \theta)$ such that the Hellinger distance reaches the minimum value in the parameter spaces $\Theta = (0, 1] \times (0, 2] \times (0, 2) \subset \mathbb{R}^3$.

$$G(\theta, \epsilon, \alpha) = \frac{1}{2} \int_{\mathbb{R}^1} \left(\sqrt{p(x, \theta, \epsilon, \alpha)} - \sqrt{p_d(x)} \right)^2 dx.$$

The values of the model parameters $\hat{\epsilon} = 0.307$, $\hat{\theta} = 1.05$ and $\hat{\alpha} = 1.7380$ are achieved by minimizing the Hellinger distance $G(\hat{\alpha}, \hat{\theta}, \hat{\epsilon}) = 0.0027$ over the parameter space $\Theta \subset \mathbb{R}^3$. Estimated results $\hat{\epsilon} = 0.307$, $\hat{\theta} = 1.05$ and $\hat{\alpha} = 1.7380$ defines a slice plane in the ϵ -axis, θ -axis, or α -axis direction as shown in Fig. 7.

4 Conclusion

In summary, we consider a non-Gaussian dynamical system containing uncertain parameters. An approach of parameter estimation is proposed by numerical optimization of the Hellinger distance between two probability distributions. The one probability density $p(x)$ is a solution of the nonlocal Fokker-Planck equation at time t for a stochastic dynamical system $X(t)$ driven by an α -stable Lévy process. The other

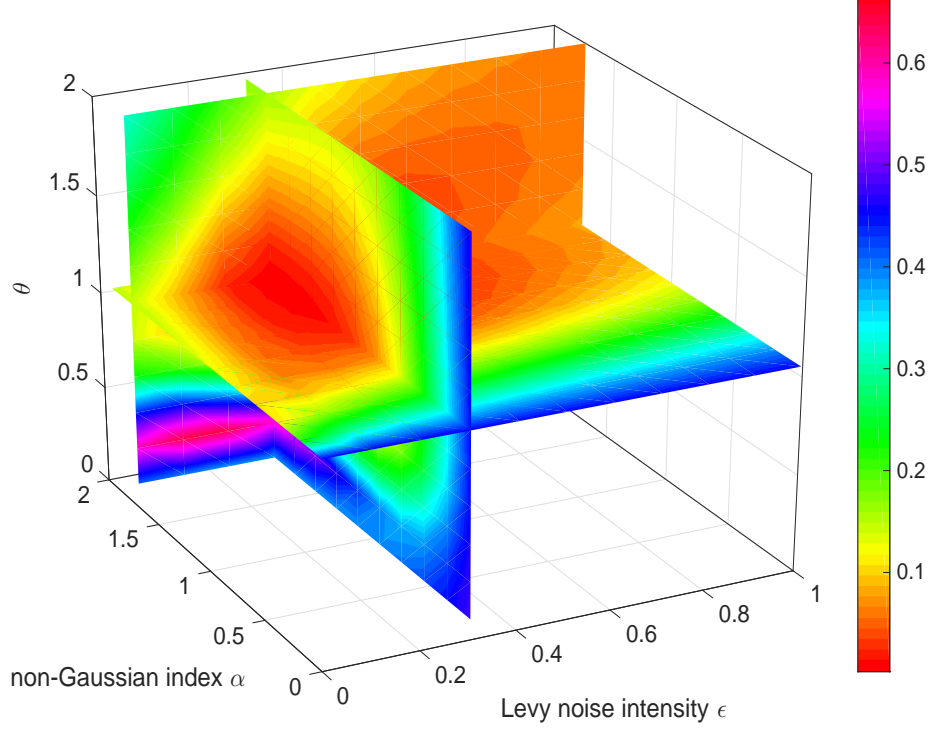


Figure 7: The Hellinger distance estimation of α , ϵ and θ . The estimated $\hat{\epsilon} = 0.307$, $\hat{\theta} = 1.05$ and $\hat{\alpha} = 1.7380$ for the minimum $G(\hat{\alpha}, \hat{\theta}, \hat{\epsilon}) = 0.0027$.

one is the empirical probability density p_d from observations data of discrete version of the process. The approach is used to find all out the estimation of single parameter and multiple parameters, by a numerical optimization of the Hellinger distance over the parameters space. The results of an example verified that this method is feasible for estimating non-Gaussian parameters α , ϵ and other system parameter by the Hellinger distance. Compared with the L^2 norm, the maximum absolute error distances and Sørensen distance, the Hellinger distance is the most effective method to estimate parameters in this example. Meanwhile, we could find an optimal interval for the estimated parameters.

This approach can be used to establish parameter estimations for a data-driven dynamical system, the observations data with jumps and heavy-tailed distribution. A very important future work will be a model study of the abrupt climate changes in the Dansgaard-Oeschger events with non-Gaussian distribution. The approach would be applied to estimate the system parameters and non-Gaussian parameters in this model.

Data Availability

All computational results are implemented with MATLAB R2015b running in an Intel Xeon(R) CPU E5-2667 v4 @ 3.20 GHz machine. The data that support the findings of this study are available in GitHub at: <https://github.com/yayun55/Estimating-uncertainty-for-the-observed-non-Gaussian-data>.

References

- [1] J. Bishwal. *Parameter estimation in stochastic differential equations*. Springer-Verlag Berlin Heidelberg, 2008.
- [2] S. Iacus. *Simulation and inference for stochastic differential equations: With R examples*. Springer-Verlag New York, 2008.
- [3] Z. Botev, J. Grotowski, and D. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5):2916–2957, 2010.
- [4] O. Papaspiliopoulos, Y. Pokern, G. Roberts, and A. Stuart. Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, 99(3):511–531, 2012.
- [5] M. Beaumont, W. Zhang, and D. Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.
- [6] C. Honisch and R. Friedrich. Estimation of Kramers-Moyal coefficients at low sampling rates. *Physical Review E*, 84(2):066701, 2011.
- [7] S. Siegert, R. Friedrich, and J. Peinke. Analysis of data sets of stochastic systems. *Physics Letters A*, 243(5-6):275–280, 1998.
- [8] P. Batz, A. Ruttor, and M. Opper. Variational estimation of the drift for stochastic differential equations from the empirical density. *Journal of Statistical Mechanics-Theory and Experiment*, 2016(8):083404, 2016.
- [9] P. Ditlevsen. Observation of α -stable noise induced millennial climate changes from an ice-core record. *Geophysical Research Letters*, 26(10):1441–1444, 1999.
- [10] L. Cai, N. Friedman, and X. Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(7082):358–362, 2006.
- [11] N. Humphries, N. Queiroz, J. Dyer, N. Pade, M. Musyl, K. Schaefer, D. Fuller, J. Brunnschweiler, T. Doyle, J. Houghton, G. Hays, C. Jones, L. Noble, V. Wearmouth, E. Southall, and D. Sims. Environmental context explains Lévy and brownian movement patterns of marine predators. *Nature*, 465(7301):1066–1069, 2010.
- [12] R. Mantegna. Lévy walks and enhanced diffusion in milan stock exchange. *Physica A: Statistical Mechanics and its Applications*, 179(2):232–242, 1991.

- [13] P. Montez, G. Thompson, and C. Kello. The role of semantic clustering in optimal memory foraging. *Cognitive science*, 39(8):1925–1939, 2015.
- [14] A. Voss, V. Lerche, U. Mertens, and J. Voss. Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic bulletin and review*, 26(3):813–832, 2019.
- [15] T. Ogihara and N. Yoshida. Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes*, 14(3):189, 2011.
- [16] H. Masuda. Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. *Electronic Journal of Statistics*, 4:525–565, 2010.
- [17] Y. Hu and H. Long. Parameter estimation for Ornstein-Uhlenbeck processes driven by α -stable Lévy motions. *Communications on Stochastic Analysis*, 1(2):175–192, 2007.
- [18] Y. Hu and H. Long. Least squares estimator for Ornstein-Uhlenbeck processes driven by α -stable motions. *Stochastic Processes and Their Applications*, 119(8):2465–2480, 2009.
- [19] V. Fasen. Statistical estimation of multivariate Ornstein-Uhlenbeck processes and applications to co-integration. *Journal of Econometrics*, 172(2):325–337, 2013.
- [20] P. Gopikrishnan, V. Plerou, L. Amaral, M. Meyer, and H. Stanley. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60(5):5305–5316, 1999.
- [21] B. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174, 1975.
- [22] J. Yang and J. Duan. Quantifying model uncertainties in complex systems. In *Stochastic Analysis with Financial Applications*, pages 221–252. Springer, 2011.
- [23] Y. Cheng, Y. Hu, and H. Long. Generalized moment estimators for α -stable Ornstein-Uhlenbeck motions from discrete observations. *Statistical Inference for Stochastic Processes*, 23(1):53–81, 2020.
- [24] T. Gao and J. Duan. Quantifying model uncertainty in dynamical systems driven by non-Gaussian Lévy stable noise with observations on mean exit time or escape probability. *Communications in Nonlinear Science and Numerical Simulation*, 39:1–6, 2016.
- [25] Z. Chen and X. Chen. Maximum likelihood estimation for symmetric α -stable Ornstein-Uhlenbeck processes. *Stochastics and Dynamics*, 2150018, 2020.
- [26] Y. Li and J. Duan. A data-driven approach for discovering stochastic dynamical systems with non-Gaussian Lévy noise. <https://arxiv.org/pdf/2005.03769.pdf>.
- [27] X. Sun, X. Li, and Y. Zheng. Governing equations for probability densities of Marcus stochastic differential equations with Lévy noise. *Stochastics and Dynamics*, 17(05):1750033, 2017.

- [28] T. Gao, J. Duan, and X. Li. Fokker-planck equations for stochastic dynamical systems with symmetric Lévy motions. *Applied Mathematics and Computation*, 278:1–20, 2016.
- [29] B. Moghaddam, A Mendes Lopes, J. Tenreiro Machado, and Z. Mostaghim. Computational scheme for solving nonlinear fractional stochastic differential equations with delay. *Stochastic Analysis and Applications*, 37(6):893–908, 2019.
- [30] F. Khane Keshi, B. Moghaddam, and A. Aghili. A numerical technique for variable-order fractional functional nonlinear dynamic systems. *International Journal of Dynamics and Control*, 7(4):1350–1357, 2019.
- [31] B. Moghaddam, Z. Mostaghim, A. Pantelous, and J. Tenreiro Machado. An integro quadratic spline-based scheme for solving nonlinear fractional stochastic differential equations with constant time delay. *Communications in Nonlinear Science and Numerical Simulation*, 92:105475.
- [32] D. Nie, J. Sun, and W. Deng. Numerical algorithm for the space-time fractional fokker–planck system with two internal states. *Numerische Mathematik*, 146(3):481–511, 2020.
- [33] D. Applebaum. *Lévy processes and stochastic calculus*. Cambridge University Press, 2009.
- [34] K. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.
- [35] X. Zhang. Densities for sdes driven by degenerate α -stable processes. *Annals of Probability*, 42(5):1885–1910, 2014.
- [36] Y. Song and X. Zhang. Regularity of density for SDEs driven by degenerate Lévy noises. *Electronic Journal of Probability*, 20:1–27, 2015.
- [37] Z. Chen, R. Song, and X. Zhang. Stochastic flows for Lévy processes with Hölder drifts. *Revista Matemática Iberoamericana*, 34(4):1755–1788, 2018.
- [38] H. Kaper and H. Engler. *Mathematics and climate*. SIAM, 2013.
- [39] P. Cessi. A simple box model of stochastically forced thermohaline flow. *Journal of physical oceanography*, 24(9):1911–1920, 1994.
- [40] R. Beran. Minimum Hellinger distance estimates for parametric models. *Annals of Statistics*, 5(3):445–463, 1977.
- [41] J. Duan. *An introduction to stochastic dynamics*. Cambridge University Press, 2015.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Y. Zheng designed the research. Y. Zheng and F. Yang performed computations and wrote the first draft of the manuscript. J. Kurths and J. Duan analysed the results and concepts development. All authors conducted research discussions and reviewed the manuscript.

Acknowledgements

We would like to thank Xiaoli Chen, Xiujun Cheng and Yang Liu for discussions about computation. This work is supported by the National Natural Science Foundation of China (grants No.11801192 and No. 11771449), Jiangsu University Project Grant (No. 20JDG071) and Russian Ministry of Science and Education "Digital biodesign and personalised healthcare".