

# Statistical Inference of 1D Persistent Nonlinear Time Series and Application to Predictions

Johannes A. Kassel\* and Holger Kantz†  
 Max Planck Institute for the Physics of Complex Systems,  
 Nöthnitzer Straße 38, 01187 Dresden, Germany, EU  
 (Dated: August 2, 2021)

We introduce a method for reconstructing macroscopic models of one-dimensional stochastic processes with long-range correlations from sparsely sampled time series by combining fractional calculus and discrete-time Langevin equations. The method is illustrated for the ARFIMA(1,d,0) process and a nonlinear auto-regressive toy model with multiplicative noise. We reconstruct a model for daily mean temperature data recorded at Potsdam (Germany) and use it to predict the first frost date by computing the mean first passage time of the reconstructed process and the 0 °C temperature line, illustrating the potential of long-memory models for predictions in the subseasonal-to-seasonal range.

*Introduction*—Predicting the dynamics of complex systems with models inferred from data has been a long-standing endeavor of science. If such models are stochastic they can capture quite naturally erratic fluctuations in the observed data. We will discuss the large body of literature on the reconstruction of Markov processes below. However, in many real world data sets, violations of Markovianity by long-range temporal correlations have been observed. For a stationary process with light-tailed increment distribution, the Hurst exponent  $H$  measures such temporal long-range correlations [1]. For  $H > 0.5$ , the process exhibits persistent long-range correlations. For  $H = 0.5$ , the process is only short-range correlated. Models for long-range correlations emerged after Hurst’s study of the reservoir capacity for the river Nile [2]. Later on, long-range correlations were found in data sets of temperature anomalies [3, 4], river runoffs [5], extreme events return intervals [6], biological systems [7, 8], and economics [9]. The earliest models generating long-range correlations are Fractional Brownian Motion (FBM) [10] in continuous time and ARFIMA processes [11, 12] in discrete time. The ARFIMA(1,d,0) process is defined as:

$$y_{t+1} = \phi y_t + (1 - B)^{-d} \xi_t, \quad (1)$$

in which the positive real number  $\phi$  is the autoregressive parameter,  $B$  is the backshift operator and  $\xi_t$  Gaussian white noise. It has the asymptotic Hurst exponent  $H = 0.5 + d$ . Figure 1 shows conditional averages of  $y_t$ ,  $E(y_t|y_0 \in [2.9995, 3.0005])$  as a function of  $t$  for various values of the memory parameter  $d$ , where the condition requires that  $y_0 \in [2.9995, 3.0005]$ . The short-range limit of this example,  $d = 0$ ,  $H = 1/2$ , is an AR(1) process and has an autocorrelation time of  $\tau = -1/\ln \phi \approx 2.3$ . The much slower relaxation of this conditional mean to the sample mean of the process (which is 0) demonstrates that memory in the noise can lead to enhanced predictability of the process. Therefore, it is beneficial to reconstruct such models from data, if there are clear indications for temporal long-range correlations, instead of ignoring them.

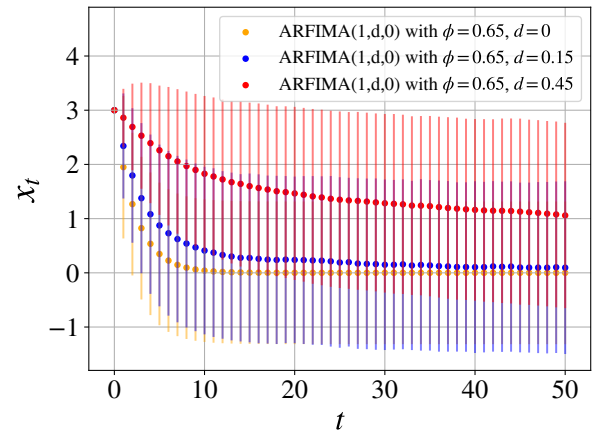


FIG. 1. Conditional averages of ARFIMA(1,d,0) processes with  $\phi = 0.65$  and various parameter values  $d$  relax to zero on different time scales. For larger  $d$ , the memory of the noise is stronger, resulting in a slower relaxation towards the mean of the process. This indicates that for processes with long-range correlations ( $d > 0$ ), prediction horizons are longer than for processes without long-range correlations. All averages are conditioned on  $y_0 \in [2.9995, 3.0005]$ . The memory length for processes with  $d > 0$  is  $M = 250$ , the sample size is  $N = 10^4$ . Error bars indicate standard deviations.

Today, there are many approaches to reconstructing stochastic models from data. Examples include Generalized Langevin equations [13, 14], Fractional Klein-Kramers equations [15], underdamped Langevin equations [16], Fokker-Planck equations [17–20], and discrete-time ARFIMA and NARMA models [21, 22]. While all of these approaches deal with either low sampling rates, long-range correlated data, nonlinear drift terms, multiplicative noise or single-trajectory data, none of them covers all of these complications for model reconstruction at once. However, in many applications e.g. geophysical time series recordings, neither trajectory ensembles nor highly sampled data sets are available, when the time series exhibit both non-trivial short-range and long-

range behavior. Király and Jánosi propose a method for the model reconstruction of daily temperature anomalies with long-range correlated input noise in an ad hoc and approximate way. [23] Here, we extend this pioneering work to a generally valid framework for the reconstruction of discrete-time models and illustrate the predictive power of long-memory models.

In the remainder of this letter, we describe our method and illustrate it by applying it to the ARFIMA(1,d,0) process, to a non-trivial toy model, and to daily mean temperature data. Finally, we use a reconstructed stochastic model of daily mean temperature anomalies to predict the first frost date in Potsdam, Germany, and assess the performance of the prediction.

*Method*– We exploit the scale freedom of long-range correlations and decompose the long-range and short-range behavior of stochastic time series. Firstly, we remove long-range correlations using the Grünwald-Letnikov fractional derivative resulting in a process which is approximately Markovian. Then, we reconstruct the short-range dynamics with a discrete-time Langevin equation. Finally, we numerically create sample paths with the inferred Langevin equation and introduce long-range temporal correlations again employing the Grünwald-Letnikov fractional integral also used in ARFIMA processes.

We start with a one-dimensional, stationary time series  $\{y_t\}_{1 \leq t \leq N}$  of length  $N$ , which exhibits an asymptotically constant Hurst exponent  $H > 0.5$ . The numerical value of  $H$  may be determined by Detrended Fluctuation Analysis (DFA) [24, 25] or other methods, among them R/S statistics [2], and Wavelet transforms [26, 27]. We use the first-order finite difference approximation of the Grünwald-Letnikov fractional derivative of order  $d = H - \frac{1}{2}$  with a finite difference of  $\Delta t = 1.0$ , defined as [28]

$${}_{t-M}D_t^d y_t = \sum_{j=0}^M \omega_j^{(d)} y_{t-j}; \quad \omega_j^{(d)} = (-1)^j \binom{d}{j}. \quad (2)$$

Here,  $M$  defines the memory length of the fractional operation which for ARFIMA or FBM is infinite. For numerical ease we use the recurrence relation  $w_j^{(d)} = (1 - \frac{d+1}{j}) w_{j-1}^{(d)}$  with  $w_0^{(d)} = 1$ . Also, for a given time series, choosing an appropriate finite  $M$  is a trade-off between the loss of  $M$  data points and the time scale of the long-range correlations to be removed. Choosing  $M = N/2$  would be optimal, but increased statistical fluctuations in the subsequent analysis advice smaller  $M$ . Removal of long-range correlations from time series using fractional calculus has been applied e.g. in [29, 30].

The values of the resulting fractionally differenced time series are denoted by  $\{{}_{t-M}D_t^d y_t\} = \{x_t\}$ , which we consider Markovian. We now model the time series

$\{x_t\}_{1 < t < N-M}$  with a discrete-time Langevin equation

$$x_{t+1} = f(x_t) + g(x_t) \xi_t. \quad (3)$$

Reminiscent of the continuous-time Langevin equation we refer to  $f(x_t)$  as drift and to  $g(x_t)$  as diffusion. Here, both  $f(x_t)$  and  $g(x_t)$  are allowed to be nonlinear resulting in a nonlinear restoring force and multiplicative noise,  $\xi_t$  denotes Gaussian white noise with  $\langle \xi_t \rangle = 0$  and  $\langle \xi_t \xi_{t'} \rangle = \delta_{tt'}$ . We assume  $g(x_t) \geq 0$  for  $x_t \in (-\infty, \infty)$ . The subsequent scheme is inspired by the reconstruction scheme for time-discrete NARMA models [21, 31]. At first, we make an ansatz  $\Phi(x_t; \phi)$ ,  $\phi = (\phi_1, \phi_2, \dots)$  for the drift  $f(x_t)$ . The functional form of  $\Phi$  requires an educated guess upon inspection of the data in the  $(x_{t+1}, x_t)$  plane. Demanding stability of the process requires  $f(x_t)$  to monotonically decrease in  $x_t$  for  $x_t \rightarrow \pm\infty$ . We then find the optimal parameters  $\hat{\phi}$  by a least-squares fit, i.e.

$$\hat{\phi} = \arg \min_{\{\phi\}} \sum_{t=1}^{N-1} (x_{t+1} - \Phi(x_t; \phi))^2 = \arg \min_{\{\phi\}} \sum_{t=1}^{N-1} R_t(\phi)^2. \quad (4)$$

For a drift function  $\Phi(x_t, \hat{\phi})$  which resembles  $f(x_t)$ , the averaged squared residual amounts to  $\langle R_t^2 \rangle = g(x_t)^2 \langle \xi_t^2 \rangle = g(x_t)^2$ , because of assumptions about the noise. Hence, we make an ansatz  $\Theta(x_t; \theta)$ ,  $\theta = (\theta_1, \theta_2, \dots)$  for the squared residuals. Again, an educated guess is needed for its functional form. Performing a least-squares fit yields the optimal parameters for approximating  $g(x_t)^2$ .

With the acquired parameters, we can generate trajectories employing the following discrete-time Langevin equation:

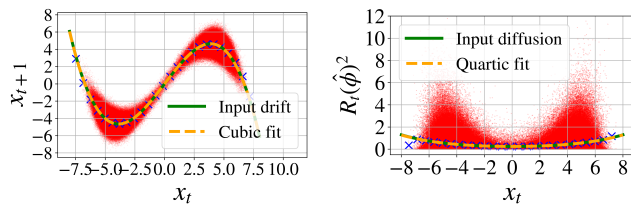
$$x_{t+1} = \Phi(x_t, \hat{\phi}) + \sqrt{\Theta(x_t, \hat{\theta})} \xi_t. \quad (5)$$

Here,  $\xi_t$  is Gaussian white noise with zero mean and variance one. By construction, time series generated using Eq. 5 are Markovian and should have similar stochastic properties as the fractionally differenced time series  $\{x_t\}$ .

Finally, we fractionally integrate the model time series, adding long-range correlations to the model data. For this purpose, we employ the first-order finite difference approximation of the Grünwald-Letnikov fractional integral which is obtained by setting  $d \rightarrow -d$  in Eq. 2 and reads:

$${}_{t-M}I_t^d x_t = \sum_{j=0}^M (-1)^j \binom{-d}{j} x_{t-j}. \quad (6)$$

Our approach neglects measurement noise. Since we are interested in reconstructing a macroscopic model possessing the same statistical properties as the original time series, we consider potential measurement noise as an indistinguishable part of the process. Choosing appropriate functions  $\Phi$  and  $\Theta$  is crucial for obtaining a suitable



(a) Drift of bimodal toy model (b) Diffusion of bimodal toy model

FIG. 2. Parameter inference for toy model defined by Eq. 7. Left panel 2(a) shows the drift inference of the model, right panel 2(b) shows the diffusion inference of the model. Red dots are the  $N = 10^6$  data points. Blue crosses show average values for 25 bins of equal width, only shown for illustration. Orange curves show the results of least-squares fits for polynomials of order three, and four, respectively. Green dashed curves show input drift and input diffusion, respectively. Orange and green curves are in perfect agreement.

model. Therefore, we advise testing various functions and base the selection both on goodness of fit as well as comparisons of model data and original data.

*ARFIMA(1,d,0) process and the discrete-time Langevin equation*– We demonstrate the two parts of our method with the ARFIMA(1,d,0) process and a toy model defined by a non-linear discrete-time Langevin equation. From the definition of the ARFIMA(1,d,0) process  $y_t$  (cf. Eq. 1), it is clear that by applying the finite difference fractional derivative (cf. Eq. 2) we obtain the AR(1) process:

$$x_{t+1} = \phi x_t + \xi_t, \quad x_t = (1 - B)^d y_t = \lim_{M \rightarrow \infty} {}_{t-M}D_t^d y_t.$$

Due to linearity, the auto-regressive parameter  $\phi$  is the same as in the ARFIMA(1,d,0) model. Hence, inference of  $\phi$  from the fractionally differenced process and subsequent fractional integration of the inferred process yields the original process here.

The following toy model process possesses a bimodal distribution and illustrates solely the second part of our method for nonlinear functions  $f(x_t)$  and  $g(x_t)$ :

$$x_{t+1} = -0.04 x_t^3 + 1.8 x_t + (0.01 x_t^2 + 0.5) \xi_t, \quad (7)$$

with  $\xi_t$  as before. We make polynomial ansatzes of order three and four for the drift  $\Phi(x_t)$  and diffusion  $\Theta(x_t)$ , respectively. Figure 2 displays model data as well as the perfect agreement of input drift and diffusion functions and their reconstructions. The reconstruction works also with a fifth order polynomial for  $\Phi(x_t)$  and a sixth order polynomial for  $\Theta(x_t)$ .

*Daily Temperature Data and First Frost Prediction*– We apply our method to daily mean 2m-temperature data of the Potsdam Telegrafenberg weather station and predict the first frost date in late autumn using the first passage time of the reconstructed process with the

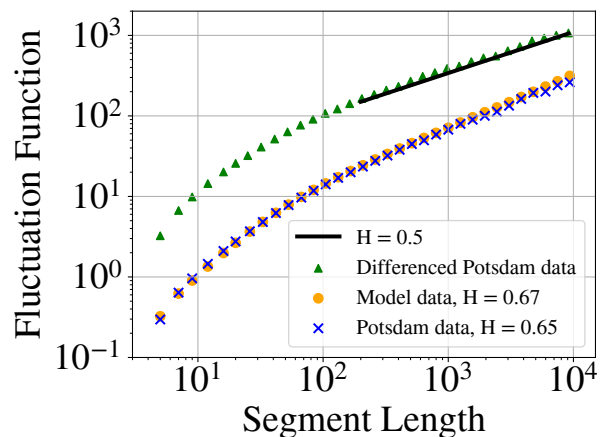


FIG. 3. Detrended Fluctuation Analysis (DFA-3) of daily mean temperature anomalies (green triangles), fractionally differenced daily mean temperature anomalies (blue crosses) and model data (orange dots). Offset for improved visibility. The asymptotic slope of the fluctuation functions  $H$  of the daily mean temperature anomalies and the model data coincide almost perfectly. The slope of the fractionally differenced daily temperature anomalies approaches the  $H = 0.5$  line, indicating the absence of long-range correlations.

zero temperature boundary. The data is provided by the European Climate Assessment & Dataset project team (<https://www.ecad.eu/>) [32]. The Potsdam temperature data set consists of an uninterrupted time series starting January 1st 1893 and is therefore apt for our analysis. Neglecting the daily temperature cycle, we consider the temperature data set as a time series of a discrete-time stochastic process with two additional trends, namely seasonal cycle (also called climatology) and climate change. We approximate the seasonal cycle by fitting a second-order Fourier series to the data, adding a quadratic function in time to account for the nonstationarity of the temperature time series due to climate change. The resulting stationary time series referred to as temperature anomalies is approximately Gaussian [33, Fig.2, p.9246]. Here, we use DFA-3 to determine the Hurst exponent resulting in  $H = 0.65$  (cf. Figure 3).

Following the recipe described above, we fractionally differentiate the temperature anomalies with  $d = H - 0.5$  and a memory length of three years ( $M = 1095$ ). Choosing longer memory ranges does not improve the model. For the drift and diffusion terms, we make a polynomial ansatz of order three and order four, respectively. Figure 4 displays the estimated drift and diffusion functions for the fractionally differenced Potsdam Telegrafenberg daily mean temperature anomalies. Király and János also report nonlinearities for drift and diffusion of temperature anomalies for an aggregate of temperature time series of 20 Hungarian weather stations. [23, Fig.3, p.4] Their data shows more pronounced nonlineari-

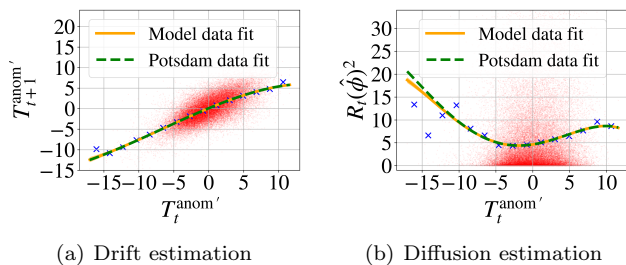


FIG. 4. Estimation of drift (cf. panel 4(a)) and diffusion (cf. panel 4(b)) of the discrete-time Langevin equation for fractionally differenced daily mean temperature anomalies of the Potsdam Telegrafenberg weather station. Red dots are the fractionally differenced anomalies (see panel 4(a)), and their squared residuals (cf. Eq. 4, panel 4(b)). The blue crosses are bin averages of the red dots, displayed for illustration only. The green curves are results of least-squares fits of polynomials of order three for the drift and order four for the diffusion. The orange curves are results of least-square fits of model data (100 samples of the length of the Potsdam data) generated with Eq. 5 and obtained parameters of the green curves. There are small deviations of the diffusion for large negative anomalies between the Potsdam data and the modal data due to the numerical stability constraint.

ties for drift and diffusion than the Potsdam temperature anomalies because of more data points for large anomalies where nonlinearities are more dominant.

To ensure numerical stability of the discrete-time Langevin equation defined by the estimated drift and diffusion functions, we set  $\Theta(x_t > x_{\max}) = \Theta(x_{\max})$  and  $\Theta(x_t < x_{\min}) = \Theta(x_{\min})$ . We then fractionally integrate a discrete-time Langevin trajectory generated with the drift and diffusion parameters obtained. Figure 5 displays the cumulative histograms, autocorrelation functions and power spectral densities of the temperature anomalies and model trajectories (see Figure 3 for the Hurst parameter estimation). They are in very good agreement.

The reconstructed process may serve for making predictions. We predict the first frost date for the Potsdam Telegrafenberg weather station by computing the first passage time distribution of generated process trajectories and the zero temperature line for a sample size of thirty years. We choose the 31st of October as the forecast start date. For each sample year, we cut the Potsdam daily mean temperature time series at the 31st of October, resulting in a time series from January 1st 1893 to the 31st of October of the sample year. After removal of the seasonal cycle, we infer model parameters with our method. Using the reconstructed model, we generate  $25 \times 10^3$  trajectories using Eq. 5, setting the fractionally differenced temperature on the forecast start date as the initial condition. We add the generated trajectory to the fractionally differenced temperature anomalies, fractionally integrate the concatenated

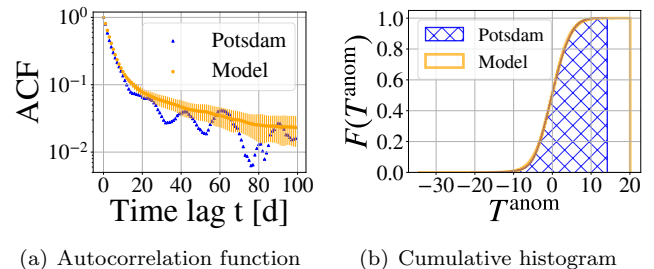


FIG. 5. Comparison of Potsdam daily mean temperature anomalies and model data. The model time series consists of 100 samples of the length of the Potsdam data set. The power spectral density is estimated with a periodogram and Welch's method. The model data shows slightly higher variance than the Potsdam data. The autocorrelation function of the Potsdam data exhibits some small-scale oscillations not explained by our model. The power spectral density of the model agrees well with the Potsdam data apart from a kink at the maximum frequency.

new trajectory, add the seasonal cycle and determine its first passage time with the  $0^\circ\text{C}$  temperature line. The mean first passage time over the ensemble of  $25 \times 10^3$  values is the predicted first frost date. For a benchmark prediction we fit a parabola to the observed frost dates of the years before the sample year, paralleling the climate change correction, and extrapolate it to the sample year. Figure 6 shows the observed first frost date, the predicted first frost date and its standard deviation, the benchmark prediction and the zero-crossing of the seasonality cycle for the years 1991 – 2020. The bias of the predicted first frost sample average amounts to  $-0.64$  days, meaning our prediction is almost unbiased. We use the root-mean-square error (RMSE) and the mean absolute error (MAE) to measure the prediction performance. The RMSE of our prediction is smaller than the variance of the observed first frost dates, indicating our prediction narrows the uncertainty of the predicted event. RMSE and MAE (cf. caption of Figure 6) show that the prediction performs much better than the seasonality but only slightly better than the benchmark estimation. We note that the variance of the observed first frost date is much larger than the variance of the prediction. In real weather, the first frost date is impacted by many factors, e.g. large-scale weather patterns not captured

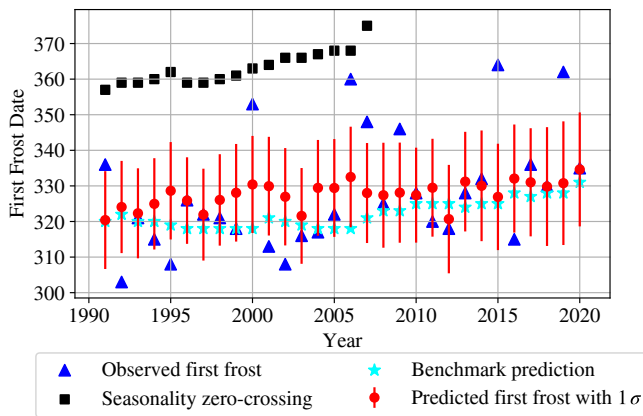


FIG. 6. First Frost Prediction Results. Dark blue triangles are the observed first frost dates of the Potsdam Telegrafenberg daily mean temperature data set. Light blue stars indicate the benchmark prediction of the first frost date obtained by fitting a parabola to the previous observed first frost dates since 1893. Black squares are the zero-crossings of the seasonality cycle for years in which they exist. Red dots are the predicted first frost date with one standard deviation of the first frost date distribution. Accuracy of estimators: Prediction: RMSE = 15.3 d and MAE = 11.5 d, benchmark prediction: RMSE = 16.8 d and MAE = 11.7 d, seasonality: RMSE = 41.8 d and MAE = 39.2 d, standard deviation of observed first frost dates:  $\sigma = 16.1$  d. The first frost prediction performs slightly better than the benchmark prediction.

by the local daily mean temperature. Commemorating we solely use a one-dimensional time series to predict an event in a high-dimensional complex system, we expect better prediction performances for reconstructed models in more-dimensional systems. Additionally, larger values of the memory parameter  $d$  would also contribute to larger prediction horizons (cf. Figure 1).

We thank Philipp G Meyer, Katja Polotzek, Christoph Streissnig, and Benjamin Walter for fruitful discussions and Steffen Peters for IT support.

\* jkassel@pks.mpg.de

† kantz@pks.mpg.de

[1] L. Chen, K. E. Bassler, J. L. McCauley, and G. H. Gunaratne, *Phys. Rev. E* **95**, 042141 (2017).  
 [2] H. E. Hurst, *Transactions of the American Society of Civil Engineers* **116**, 770–799 (1951).  
 [3] K. Fraedrich and R. Blender, *Phys. Rev. Lett.* **90**, 108501 (2003).  
 [4] J. F. Eichner, E. Koscielny-Bunde, A. Bunde, S. Havlin, and H.-J. Schellnhuber, *Phys. Rev. E* **68**, 046133 (2003).  
 [5] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin, *Journal of Geophysical Research: Atmospheres* **111** (2006).

[6] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, *Phys. Rev. Lett.* **94**, 048701 (2005).  
 [7] K. Y. Wan and R. E. Goldstein, *Phys. Rev. Lett.* **113**, 238103 (2014).  
 [8] J. Echeverria, M. Woolfson, J. Crowe, B. Hayes-Gill, G. Croaker, and H. Vyas, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **13**, 467 (2003).  
 [9] R. T. Baillie, *Journal of Econometrics* **73**, 5 (1996).  
 [10] B. B. Mandelbrot and J. W. Van Ness, *SIAM Review* **10**, 422 (1968).  
 [11] C. W. J. Granger and R. Joyeux, *Journal of Time Series Analysis* **1**, 15 (1980).  
 [12] J. R. M. Hosking, *Biometrika* **68**, 165 (1981).  
 [13] S. C. Kou and X. S. Xie, *Phys. Rev. Lett.* **93**, 180603 (2004).  
 [14] H. Lei, N. A. Baker, and X. Li, *Proceedings of the National Academy of Sciences* **113**, 14183 (2016).  
 [15] P. Dieterich, R. Klages, R. Preuss, and A. Schwab, *Proceedings of the National Academy of Sciences* **105**, 459 (2008).  
 [16] D. B. Brückner, P. Ronceray, and C. P. Broedersz, *Phys. Rev. Lett.* **125**, 058103 (2020).  
 [17] C. Honisch, R. Friedrich, F. Hörner, and C. Denz, *Phys. Rev. E* **86**, 026702 (2012).  
 [18] M. Ragwitz and H. Kantz, *Phys. Rev. Lett.* **87**, 254501 (2001).  
 [19] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase, *Phys. Rev. Lett.* **97**, 090603 (2006).  
 [20] M. Tabar, *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems: Using the Methods of Stochastic Processes*, Understanding Complex Systems (Springer International Publishing, 2019).  
 [21] A. J. Chorin and F. Lu, *Proceedings of the National Academy of Sciences* **112**, 9804 (2015).  
 [22] T. Graves, R. B. Gramacy, C. L. E. Franzke, and N. W. Watkins, *Nonlinear Processes in Geophysics* **22**, 679 (2015).  
 [23] A. Király and I. M. Jánosi, *Phys. Rev. E* **65**, 051102 (2002).  
 [24] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, *Phys. Rev. E* **49**, 1685 (1994).  
 [25] M. Höll, K. Kiyono, and H. Kantz, *Phys. Rev. E* **99**, 033305 (2019).  
 [26] I. Simonsen, A. Hansen, and O. M. Nes, *Phys. Rev. E* **58**, 2779 (1998).  
 [27] P. Abry and D. Veitch, *IEEE Trans. Inf. Theory* **44**, 2 (1998).  
 [28] I. Podlubny, *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications* (Elsevier Science, 1998).  
 [29] I. Petráš and J. Terpák, *Mathematics* **7** (2019).  
 [30] N. Yuan, Z. Fu, and S. Liu, *Journal of Geophysical Research: Atmospheres* **118**, 12962 (2013).  
 [31] K. K. L. Fei Lu and A. J. Chorin, *Communications of Applied Mathematics and Computational Science* **11**, 187 (2016).  
 [32] A. M. G. Klein Tank *et al.*, *International Journal of Climatology* **22**, 1441 (2002).  
 [33] M. Massah and H. Kantz, *Geophysical Research Letters* **43**, 9243 (2016).