Statistical inference of one-dimensional 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, European Union

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 autoregressive 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 $0\,°\mathrm{C}$ temperature line, illustrating the potential of long-memory models for predictions in the subseasonal-to-seasonal range.

I. INTRODUCTION

Predicting the dynamics of complex systems with models inferred from data has been a longstanding 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 [1,2]. For $H > 0.5$, the process exhibits persistent long-range correlations. For short-range correlated processes, in particular Markov processes, there exists a characteristic timescale, i.e., a minimal time separation required between two states of the process to be considered independent. Hence the process possesses no asymptotic self-similarity, resulting in $H = 0.5$ [3–5]. Models for long-range correlations emerged after Hurst's study of the reservoir capacity for the river Nile [6]. Later on, long-range correlations were found in data sets of temperature anomalies [7,8], river runoffs [9], extreme events return intervals [10], biological systems [11,12], and economics [13]. The earliest models generating long-range correlations are fractional Brownian motion (FBM) [3] in continuous time and autoregressive fractionally integrated moving average (ARFIMA) processes [14,15] in discrete time. The ARFIMA(1,d,0) pro-

cess is defined as

$$
\begin{aligned}
y_{t+1} &= \phi\, y_t + (1 - B)^{-d}\, \xi_t \\
&= \phi\, y_t + \lim_{M \to \infty} \sum_{j=0}^{M} \frac{\Gamma(j+d)}{\Gamma(j+1)\,\Gamma(d)}\, \xi_{t-j},
\end{aligned} \tag{1}
$$

in which the positive real number ϕ is the autoregressive parameter, B the backshift operator, Γ the gamma function, and ξ_t Gaussian white noise. It has the asymptotic Hurst exponent $H = 0.5 + d$ and, as Eq. (1) shows explicitly, it is not Markovian. 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 $y_0 \in [2.9995, 3.0005]$, and $E(\cdot)$ denotes the expectation value. The short-range limit of this example, $d = 0, H = 1/2$, is the Markovian $AR(1)$ process and has an autocorrelation time of $\tau = -1/\ln\phi \approx 2.3$. The much faster 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.

Today, there are many approaches to reconstructing stochastic models from data. Examples include generalized Langevin equations [16,17], fractional Klein-Kramers equations [18], underdamped Langevin equations [19], Fokker-Planck equations [20–23], and discrete-time ARFIMA [24] and nonlinear autoregressive moving average (NARMA) [25] models. 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 nontrivial short-range and long-range behavior. Király

---

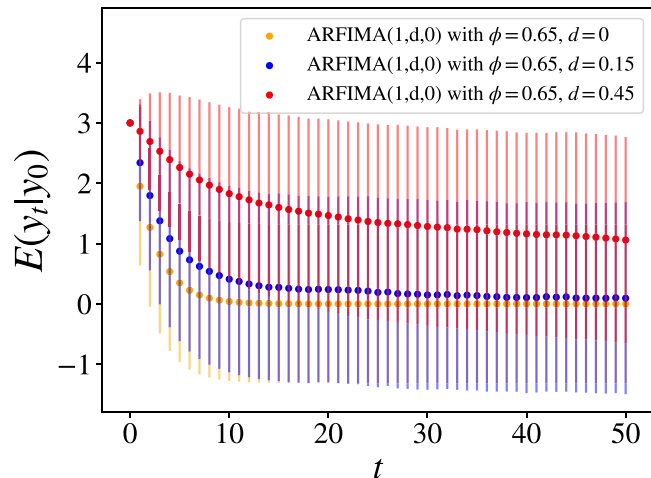[*]jkassel@pks.mpg.de

[†]kantz@pks.mpg.de

FIG. 1. Conditional averages of ARFIMA(1,d,0) processes with $\phi = 0.65$ and parameter values $d \in \{0.0, 0.15, 0.45\}$ relax to zero on different timescales. For $d = 0.0$, the process simplifies to the Markovian $AR(1)$ process (yellow curve: Analytical). The displayed curves for $d \neq 0$ are ensemble averages with $N = 10^4$ independent samples, conditioned on $y_0 \in [2.9995, 3.0005]$ and a finite memory length $M = 250$, truncating the noise integration [cf. Eq. (1)]. Error bars indicate standard deviations. 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.

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 [26]. 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 article, we describe our method and illustrate it by applying it to the ARFIMA(1,d,0) process, to a nontrivial 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.

II. METHOD

We exploit the scale freedom of long-range correlations and decompose the long-range and short-range behavior of stochastic time series. First, 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 $\{y_t\}_{1 \leqslant t \leqslant N}$ 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) [27,28] or other methods, among them R/S statistics [6], and wavelet transforms [29,30]. We use the first-order finite-difference approximation of the Grünwald-Letnikov fractional derivative of the order of $d = H - \frac{1}{2}$ with a finite difference of $\Delta t = 1.0$, defined as [31]

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

Here, M defines the memory length of the fractional operation. In theory, M goes to infinity for fractional processes [cf. Eq. (1)]. In applications, choosing an appropriate finite M is a trade-off between the loss of M data points and the timescale of the long-range correlations to be removed. Choosing $M = N/2$ would be optimal, but increased statistical fluctuations in the subsequent analysis advise smaller M. The removal of long-range correlations of time series using fractional calculus has been applied, e.g., in [32,33]. For numerical ease, we use the recurrence relation $w_j^{(d)} = (1 - \frac{d+1}{j}) \, w_{j-1}^{(d)}$ with $w_0^{(d)} = 1$ for the computation of the coefficients in Eq. (2).

The values of the resulting fractionally differenced time series are denoted by $\{_{t-M}D_t^d \, y_t\} = \{x_t\}$, which we consider approximately Markovian. We now model the time series $\{x_t\}_{1 < t < N-M}$ with a stochastic difference equation [34] and call it the 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; ξ_t denotes Gaussian white noise with $\langle \xi_t \rangle = 0$ and $\langle \xi_t \xi_{t'} \rangle = \delta_{tt'}$. We assume $g(x_t) \geqslant 0$ for $x_t \in (-\infty, \infty)$. The subsequent scheme is inspired by the reconstruction scheme for time-discrete NARMA models [25,35]. At first, we make an *Ansatz* $\Phi(x_t; \lambda)$, $\lambda = (\lambda_1, \lambda_2, \dots)$ for the drift $f(x_t)$. The functional form of Φ 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 $x_t \to \pm\infty$. We then find the optimal parameters $\hat{\lambda}$ by a least-squares fit, i.e.,

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

For a drift function $\Phi(x_t, \hat{\lambda})$ 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:

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

Here, ξ_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 $d \to -d$ in Eq. (2) and reads

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

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 Φ and Θ is crucial for obtaining a suitable 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. The assumed Markovianity of the fractionally differenced data should be tested in applications. If it is not satisfied, the discrete-time Langevin equation presented here must be replaced by a higher-order Markovian model incorporating more than one previous realization of the process.

III. 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 nonlinear 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:

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

Due to linearity, the autoregressive parameter ϕ is the same as in the ARFIMA(1,d,0) model. Hence, inference of ϕ 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 $g(x_t)$,

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

with ξ_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 also works with a fifth-order polynomial for $\Phi(x_t)$ and a sixth-order polynomial for $\Theta(x_t)$.

IV. 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 zero-temperature boundary.
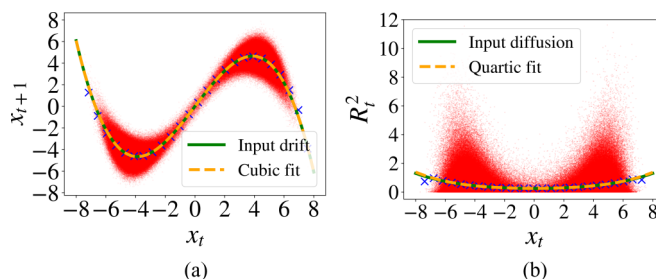

FIG. 2. Parameter inference for toy model defined by Eq. (7). (a) The drift inference of the model. (b) The diffusion inference of the model. Red dots are the $N = 10^6$ data points. Blue crosses show average values for 25 bins of equal width, only shown for illustration. Orange curves show the results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

The data are provided by the European Climate Assessment and Dataset project team [36]. The Potsdam temperature data set consists of an uninterrupted time series starting on January 1, 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 [37, Fig. 2, p. 9246]. Here, we use DFA-3, in which a cubic polynomial is used for the detrending procedure [28], to determine the Hurst exponent resulting in $H = 0.65$ (cf. Fig. 3).
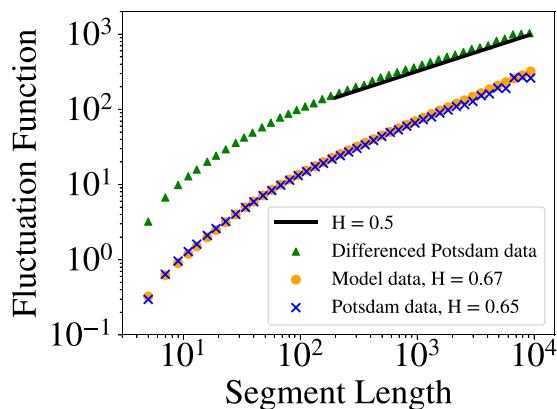

FIG. 3. Detrended fluctuation analysis (DFA-3) of daily mean temperature anomalies (green triangles), fractionally differenced daily mean temperature anomalies (blue crosses), and model data (orange dots). Offset for improved visibility. The asymptotic slope of the fluctuation functions H of the daily 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.
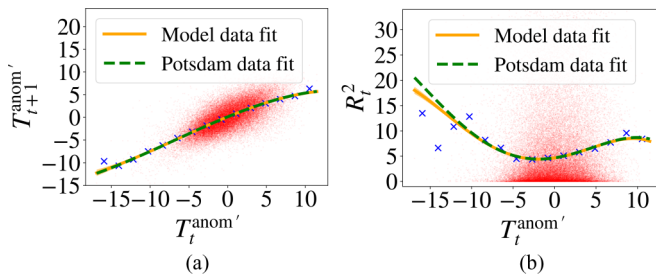
FIG. 4. Estimation of (a) drift and (b) diffusion of the discrete-time Langevin equation for fractionally differenced daily mean temperature anomalies of the Potsdam Telegrafenberg weather station. Red dots are the fractionally differenced anomalies [see (a)] and their squared residuals [cf. Eq. (4); see (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 the polynomials of order three for the drift and order four for the diffusion. The orange curves are results of least-square fits of the 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.

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. The approximate Markovianity of the fractionally differenced data is indicated by its Hurst exponent (cf. Fig. 3), the exponential decay of its autocorrelation function [cf. Fig. 5(a)], and an inspection of the dependence of the residuals $R_t(X_{t-2}, X_{t-3}|\hat{\lambda})$ on previous realizations of the process, which is negligible.

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 Jánosi also report nonlinearities for the drift and diffusion of temperature anomalies for an aggregate of temperature time series of 20 Hungarian weather stations [26, Fig. 3, p. 4]. Their data show more pronounced nonlinearities 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 obtained drift and diffusion parameters. Figure 5 displays the cumulative histograms, autocorrelation functions, and power spectral densities of the temperature anomalies and model trajectories (see Fig. 3 for the Hurst parameter estimation). They are in very good agreement, which is also confirmed by visual inspection of model time series, cf. Fig. 5(d) for one sample.

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 50 years. We choose
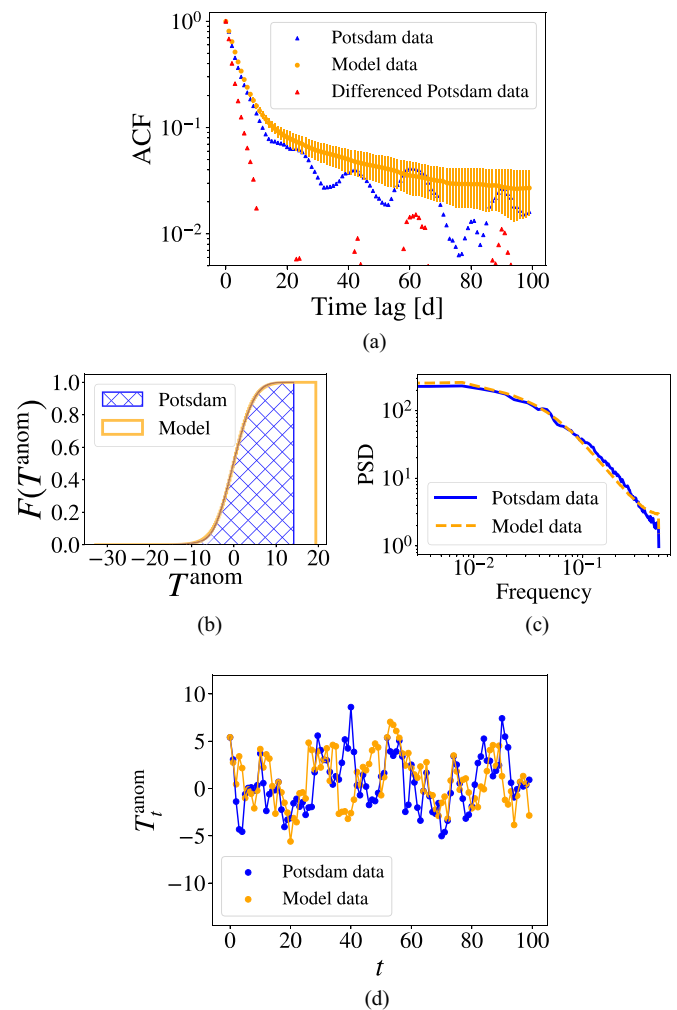

FIG. 5. Comparison of Potsdam daily mean temperature anomalies and model data. (a) The autocorrelation function of the Potsdam data exhibits some small-scale oscillations not explained by our model. The exponential decay of the fractionally differenced Potsdam data is clearly visible, indicating the approximate Markovianity of the data. (b) Cumulative distribution function. The model data show slightly higher variance than the Potsdam data. (c) The power spectral densities are estimated with a periodogram and Welch's method. Power spectral densities of the model data and the Potsdam data agree well, apart from a kink at the maximum frequency. For (a)–(c), the model data consist of 100 samples of the length of the Potsdam data set. (d) 100 data points of the Potsdam daily mean temperature anomalies and one model trajectory conditioned on the past $M = 1095$ realizations of the Potsdam daily mean temperature anomalies. Lines between data points are plotted for illustration only.

the 23rd of October as the forecast start date. For each sample year, we cut the Potsdam daily mean temperature time series at the 22nd of October, resulting in a time series from January 1, 1893 to the 22nd 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 25×10^3 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
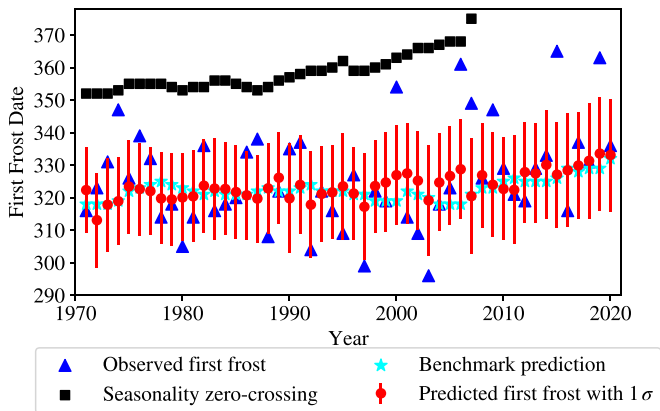
FIG. 6. First-frost prediction results. Dark blue triangles are the observed first-frost dates of the Potsdam Telegrafenberg daily mean temperature data set. Light blue stars indicate the benchmark prediction of the first-frost date obtained by fitting a parabola to the previously observed first-frost 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 = 14.7 d and MAE = 11.4 d; benchmark prediction: RMSE = 15.9 d and MAE = 11.9 d; seasonality: RMSE = 38.0 d and MAE = 34.8 d; standard deviation of observed first-frost dates: $\sigma = 15.6$ d. The first-frost prediction performs slightly better than the benchmark prediction.

concatenated new trajectory, add the seasonal cycle, and determine its first passage time with the $0\,°$C temperature line. The mean first passage time over the ensemble of 25×10^3 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 1971–2020. The bias of the predicted first-frost sample average amounts to -2.9 days, meaning our prediction only has a marginal bias compared to the average lead time of 32 days. 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. The RMSE and MAE (cf. caption of Fig. 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 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. Reconstructing these in multivariate models using the method presented in this article is part of future research. Additionally, larger values of the memory parameter d would also contribute to larger prediction horizons (cf. Fig. 1). In meteorology, the first-frost date is defined as the first passage time of the daily minimal temperature and the zero-degree temperature line, whereas we use daily mean temperature data for our analysis. The first-frost prediction results for the Potsdam minimal temperature time series are qualitatively identical, but the reconstruction of drift and diffusion is less satisfactory due to their more complex shape.

V. CONCLUSION

In this article, we propose a method for the reconstruction of one-dimensional nonlinear stochastic processes from persistent sparsely sampled time series using fractional calculus and discrete-time Langevin equations. The method performs well for ARFIMA(1,d,0) and Potsdam daily mean temperature data. A first-frost prediction for Potsdam daily mean temperature data shows predictive power to some extent.

[1] B. B. Mandelbrot and J. R. Wallis, Noah, Joseph, and operational hydrology, Water Resour. Res. **4**, 909 (1968).

[2] L. Chen, K. E. Bassler, J. L. McCauley, and G. H. Gunaratne, Anomalous scaling of stochastic processes and the Moses effect, Phys. Rev. E **95**, 042141 (2017).

[3] B. B. Mandelbrot and J. W. Van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. **10**, 422 (1968).

[4] B. B. Mandelbrot and J. R. Wallis, Some long-run properties of geophysical records, Water Resour. Res. **5**, 321 (1969).

[5] N. W. Watkins, Mandelbrot's stochastic time series models, Earth Space Sci. **6**, 2044 (2019).

[6] H. E. Hurst, Long-term storage capacity of reservoirs, Trans. Am. Soc. Civ. Eng. **116**, 770 (1951).

[7] K. Fraedrich and R. Blender, Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models, Phys. Rev. Lett. **90**, 108501 (2003).

[8] J. F. Eichner, E. Koscielny-Bunde, A. Bunde, S. Havlin, and H.-J. Schellnhuber, Power-law persistence and trends in the atmosphere: A detailed study of long temperature records, Phys. Rev. E **68**, 046133 (2003).

[9] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin, Long-term persistence and multifractality of precipitation and river runoff records, J. Geophys. Res.: Atmos. **111**, D01106 (2006).

[10] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, Long-Term Memory: A Natural Mechanism for the Clustering of Extreme Events and Anomalous Residual Times in Climate Records, Phys. Rev. Lett. **94**, 048701 (2005).

[11] K. Y. Wan and R. E. Goldstein, Rhythmicity, Recurrence, and Recovery of Flagellar Beating, Phys. Rev. Lett. **113**, 238103 (2014).

[12] J. Echeverría, M. Woolfson, J. Crowe, B. Hayes-Gill, G. Croaker, and H. Vyas, Interpretation of heart rate variability via detrended fluctuation analysis and $\alpha\beta$ filter, Chaos: Interdiscipl. J. Nonlinear Sci. **13**, 467 (2003).

[13] R. T. Baillie, Long memory processes and fractional integration in econometrics, J. Econometr. **73**, 5 (1996).

[14] C. W. J. Granger and R. Joyeux, An introduction to long-memory time series models and fractional differencing, J. Time Ser. Anal. **1**, 15 (1980).

[15] J. R. M. Hosking, Fractional differencing, Biometrika **68**, 165 (1981).

[16] S. C. Kou and X. S. Xie, Generalized Langevin Equation with Fractional Gaussian Noise: Subdiffusion within a Single Protein Molecule, Phys. Rev. Lett. **93**, 180603 (2004).

[17] H. Lei, N. A. Baker, and X. Li, Data-driven parameterization of the generalized Langevin equation, Proc. Natl. Acad. Sci. **113**, 14183 (2016).

[18] P. Dieterich, R. Klages, R. Preuss, and A. Schwab, Anomalous dynamics of cell migration, Proc. Natl. Acad. Sci. USA **105**, 459 (2008).

[19] D. B. Brückner, P. Ronceray, and C. P. Broedersz, Inferring the Dynamics of Underdamped Stochastic Systems, Phys. Rev. Lett. **125**, 058103 (2020).

[20] C. Honisch, R. Friedrich, F. Hörner, and C. Denz, Extended Kramers-Moyal analysis applied to optical trapping, Phys. Rev. E **86**, 026702 (2012).

[21] M. Ragwitz and H. Kantz, Indispensable Finite Time Corrections for Fokker-Planck Equations from Time Series Data, Phys. Rev. Lett. **87**, 254501 (2001).

[22] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase, Reconstruction of Complex Dynamical Systems Affected by Strong Measurement Noise, Phys. Rev. Lett. **97**, 090603 (2006).

[23] M. Tabar, *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems: Using the Methods of Stochastic Processes*, Understanding Complex Systems (Springer Nature Switzerland AG, Cham, Switzerland, 2019).

[24] T. Graves, R. B. Gramacy, C. L. E. Franzke, and N. W. Watkins, Efficient Bayesian inference for natural time series using ARFIMA processes, Nonlin. Proc. Geophys. **22**, 679 (2015).

[25] A. J. Chorin and F. Lu, Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics, Proc. Natl. Acad. Sci. USA **112**, 9804 (2015).

[26] A. Király and I. M. Jánosi, Stochastic modeling of daily temperature fluctuations, Phys. Rev. E **65**, 051102 (2002).

[27] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E **49**, 1685 (1994).

[28] M. Höll, K. Kiyono, and H. Kantz, Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average, Phys. Rev. E **99**, 033305 (2019).

[29] I. Simonsen, A. Hansen, and O. M. Nes, Determination of the Hurst exponent by use of wavelet transforms, Phys. Rev. E **58**, 2779 (1998).

[30] P. Abry and D. Veitch, Wavelet analysis of long-range-dependent traffic, IEEE Trans. Inf. Theory **44**, 2 (1998).

[31] I. Podlubny, *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications* (Academic Press, San Diego, California, USA, 1998).

[32] I. Petráš and J. Terpák, Fractional calculus as a simple tool for modeling and analysis of long memory process in industry, Mathematics **7**, 511 (2019).

[33] N. Yuan, Z. Fu, and S. Liu, Long-term memory in climate variability: A new look based on fractional integral techniques, J. Geophys. Res.: Atmos. **118**, 12,962 (2013).

[34] H. Tong, *Non-linear Time Series: A Dynamical System Approach*, Oxford Statistical Science Series (Clarendon Press, Oxford, UK, 1993).

[35] K. K. L. Fei Lu and A. J. Chorin, Comparison of continuous and discrete-time data-based modeling for hypoelliptic systems, Commun. Appl. Math. Comput. Sci. **11**, 187 (2016).

[36] A. M. G. Klein Tank *et al.*, Daily dataset of 20th-century surface air temperature and precipitation series for the european climate assessment, Intl. J. Climatol. **22**, 1441 (2002); also, see https://www.ecad.eu/.

[37] M. Massah and H. Kantz, Confidence intervals for time averages in the presence of long-range correlations, a case study on earth surface temperature anomalies, Geophys. Res. Lett. **43**, 9243 (2016).