Unraveling the decay of the number of unobserved ordinal patterns in noisy chaotic dynamics

Felipe Olivares,¹,* Luciano Zunino,²,³ Miguel C. Soriano,⁴ and Darío G. Pérez⁴

¹Instituto de Física, Pontificia Universidad Católica de Valparaíso (PUCV), 23-40025, Valparaíso, Chile
²Centro de Investigaciones Ópticas (CONICET La Plata - CIC), C.C. 3, 1897 Gornet, Argentina
³Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata (UNLP), 1900 La Plata, Argentina
⁴Instituto de Física Interdisciplinar y Sistemas Complejos CSIC-UIB, Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

DOI: 10.1103/PhysRevE.100.042215

I. INTRODUCTION

Permutation analysis maps a raw time series into a corresponding sequence of ordinal patterns [1]. One particular property of this symbolic mapping stands out, i.e., the emergence of the so-called forbidden patterns [2–4]. Amigó and coworkers have conclusively demonstrated, for deterministic one-dimensional maps, that not all possible ordinal patterns can be effectively materialized into orbits, which in a sense make them forbidden. Later on, Rosso et al. [5] have numerically shown that even when the presence of forbidden patterns is a characteristic of chaotic dynamics, a minimum pattern length is necessary in order to detect them. Moreover, numerical investigations suggest that forbidden ordinal patterns are also characteristic of (discrete sampled) continuous chaotic dynamics, being an effective measure of determinism when dealing with irregular and undersampled time series [6], highly irregular sampling [7], and missing data and timing jitter [8].

Conversely, for stochastic processes, all ordinal patterns will eventually appear for a sufficiently long data set. Some patterns, however, might not appear due to statistical limitations, i.e., finite lengths. On top of that, correlated stochastic processes may require even longer time series to observe all possible ordinal patterns since their occurrence depends not only on the data length but also on the degree of correlations [9,10]. Within this context, unobserved ordinal patterns are often being referred as missing.

The omnipresence of observational noise is inherent to most of the measured data sequences. Therefore, the coexistence of underlying deterministic and stochastic natures is always present in any finite chaotic measured data. In this type of data, there will exist a set of unobserved ordinal patterns, which could classify as either forbidden or missing ordinal patterns. Premised on the fact that forbidden ordinal patterns possess dynamical robustness against additive noise, together with the evidence that finite data produces missing ordinal patterns, Amigó et al. [4] introduced a practical approach to discriminate whether the underlying dynamics is uncorrelated or possess a deterministic component. This discrimination can be done by comparing the decay of the number of unobserved ordinal patterns as a function of the time series length with the decay corresponding to a white noise sequence with the same length. If the former is higher than the latter, it can be inferred that the measured data has a deterministic component. Even when successful applications of this principle in different areas of science can be found in the literature [11–16], a general mathematical model of the decay of the unobserved ordinal patterns as a function of the data length for noisy chaotic dynamics is still lacking. On the one hand, for totally uncorrelated time series, the number of missing ordinal patterns decays exponentially with the time series length [4], but on the other hand, for correlated stochastic data, it has been numerically concluded that the decay can be well fitted by a stretched exponential model [10]. A model to describe the decay of the number of unobserved ordinal patterns as a function of the time series length for chaotic dynamics in a noisy environment is needed to obtain a more robust characterization through the Amigó’s approach. The aim here is to numerically demonstrate that the stretched exponential model fits this decay for noisy chaotic dynamics in both discrete and continuous systems. Moreover, it is also shown that the fitting parameters of this model can be used for...
distinguishing between correlated stochastic and noisy chaotic dynamics.

II. ORDINAL PATTERNS

Given a one-dimensional time series (TS), \( \mathcal{X} = \{x_i; t = 1, \ldots, M\} \), first one has to choose two parameters: the order of the permutation symbols \( D \geq 2 \) (\( D \in \mathbb{N} \), the pattern length) and the lag \( \tau \) (\( \tau \in \mathbb{N} \), the time separation between the values). Subsequently, the TS is then mapped into subsets of length \( D \) of consecutive (\( \tau = 1 \)) or nonconsecutive (\( \tau > 1 \)) values, generated by (\( t \) = \( \pi \{ x_t, x_{t+\tau}, \ldots, x_{t+(D-2\tau)}, x_{t+(D-1)\tau}\} \), which assigns to each time \( t \) the \( D \)-dimensional vector of values at times \( t, t+\tau, \ldots, t+(D-1)\tau \). Clearly, more temporal information is incorporated into the vectors as the \( D \) value increases. Then, each element of the vector from zero to \( D - 1 \) is replaced by a number related to their original temporal position in the vector. By the ordinal pattern related to the time \( t \) we mean the permutation \( \pi_t \) of the values \( \{0, 1, \ldots, D - 1\} \), in accordance with the relative amplitude (strength) of each element in the ordered vector from low to high [17]. Equal values in the TS are usually ranked according to their temporal order. This is justified if the values of \( \mathcal{X} \) have a continuous distribution so that equal values are very unusual.

The value of \( D \) determines the number of accessible states, given by \( D^D \), and also conditions the minimum acceptable length, \( M \gg D! \) of the TS that one needs in order to work with a reliable statistics [1]. Bandt and Pompe suggest in their cornerstone paper to work with \( 3 \leq D \leq 7 \). Often, the lag is set equal to one. Nevertheless, other values of \( \tau \) might provide additional information. It has been shown that this parameter is strongly related, when it is relevant, to the intrinsic time scales of the system under analysis [17–20]. Ordinal symbolic mapping procedure does not require the precise optimal reconstruction of the phase space that is necessary for estimating other quantifiers of chaotic signals [21]. Consequently, the pattern length \( D \) and the lag \( \tau \) do not need to be selected following the methodologies usually employed in a conventional phase space reconstruction (the first zero of the autocorrelation function, the first minimum of the average mutual information, the false nearest-neighbor algorithm, etc.) [22].

Decay of the number of unobserved ordinal patterns

We call \( \langle N(w, D) \rangle \) the number of unobserved ordinal patterns of length \( D \) in a sequence of \( w \) data values. We proceed by (i) counting the number of unobserved ordinal patterns in sliding, nonoverlapping windows of size \( w \) along the TS of length \( M \), (ii) computing the average \( \langle N(w, D) \rangle \) over the \( \lceil M/w \rceil \) windows \( \lceil c \rceil \) is the largest integer less than or equal to the element \( c \) for each data sequence length \( w \), (iii) modeling the decay of \( \langle N(w, D) \rangle \) as a function of \( w \) with a stretched exponential function

\[
\langle N(w, D) \rangle = A e^{-Bw^\gamma},
\]

where \( A \) is a constant, \( B \) is the characteristic decay rate and \( \gamma \) is the stretching exponent—note that for \( \gamma = 1 \) a pure exponential decay is recovered. Finally, (iv) estimating the characteristic decay rate \( B \) and the stretching exponent \( \gamma \) by fitting Eq. (1) to \( \langle N(w, D) \rangle \) using a standard least-squares method.

For large window sizes, \( \langle N(w, D) \rangle \) becomes statistically unreliable since the number of windows, \( \lfloor M/w \rfloor \), is too small for an appropriate averaging procedure. Particularly, for the case of small values of \( D \), say 3, 4, and 5, \( \langle N(w, D) \rangle \) decays very fast, since the amount of possible ordinal patterns is not significantly large, for instance up to 120 for \( D = 5 \). Therefore, for small values of \( D \), the maximum window length is set such that a nonzero number of unobserved ordinal patterns is found.

III. NUMERICAL RESULTS

For the systems considered in this section we have generated TS of \( 2 \times 10^5 \) data samples. The first \( 10^5 \) iterations were discarded to avoid any transient behavior, then the effective TS length is set \( M = 10^5 \). In order to model the decay of the number of unobserved ordinal patterns as a function of the sequence length for noisy chaotic environments, white Gaussian noise of zero mean was added to the original noise-free chaotic TS considered in the following analysis. Different noise levels (NL), defined by the ratio between the standard deviation of the noise and the original signal, were considered.

For the unobserved ordinal patterns decay analysis, 100 equally spaced values for the window size \( w \in [10, w_{\text{max}}] \) were considered. The maximum window size depends on the pattern length. Specifically, we considered \( w_{\text{max}} = 110, 500, 2000, \) and 8000, for \( D = 4, 5, 6, \) and 7, respectively.

A. Logistic map

As a paradigmatic example, the logistic map is usually employed as a testing ground to illustrate new concepts in the treatment of dynamical systems. It is a polynomial mapping of degree two [23], described by (the ecologically motivated) first-order differential equation

\[
x_{n+1} = r x_n (1 - x_n),
\]

with \( 0 \leq x_0 \leq 1 \) and \( r \) the control parameter. Here, we use on \( r = 4 \), as this map presents a totally developed chaotic dynamics. Ten independent TS were generated starting from different random initial conditions. Figure 1 shows the decay of the normalized averaged \( \langle N \rangle \) as a function of the sequence length \( w \) for \( D = 6 \) and \( \tau = 1 \). In the absence of observational noise the number of unobserved ordinal patterns quickly converges to the number of forbidden patterns (for \( D = 6 \), it is 645)—see open blue circles. In contrast, when the TS is contaminated with observational noise, some of these true forbidden ordinal patterns, unrealizable by the pure chaotic dynamics, are observed in the TS as a consequence of the noise contamination. Moreover, as the noise level increases, a faster decay of the number of unobserved ordinal patterns can be observed, and it can be modeled by a stretched exponential as can be seen in Fig. 1—solid black lines. Note the excellent agreement between the simulations and the fitting independently of the noise level. Qualitative similar results were also obtained for other pattern length (\( D = 4, 5, \) and 7).

Figure 2 shows the values estimated for [Fig. 2(a)] the stretching exponent \( \gamma \) and [Fig. 2(b)] the characteristic
decay rate, $B$, associated with the stretched exponential model [Eq. (1)] as a function of the noise level. Both parameters were estimated by implementing a least-squares fit. The stretching exponent increases with the noise level for all $D$ values. This indicates that as the noise contamination increases the decay of $\langle N \rangle$ approaches an exponential decay ($y = 1$) associated with an uncorrelated TS [2,3]. Otherwise, the less noise contamination there is, the more stretched the decay becomes. Meanwhile, the characteristic decay rate decreases with the noise level and its value depends on $D$.

**B. Lorenz equations**

It is well known that the analysis of continuous systems strongly depends on the time scale used for their characterization. Stochastic or chaotic dynamics could be ascribed to the same continuous system when analyzed at different sampling frequencies. A multiscale analysis is often a practical alternative for shedding some light on this subtle issue. With the aim to illustrate the role that the sampling time has on the decay of the number of unobserved ordinal patterns, we consider the Lorenz equations given by

$$\begin{align*}
\dot{x} &= \sigma (y - x), \\
\dot{y} &= \rho x - y - xz, \\
\dot{z} &= xy - \beta z,
\end{align*}$$

(3)

with $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ and initial conditions $(x_0, y_0, z_0) = (1, 5, 10)$. The equations were numerically integrated by using a fourth-order Runge-Kutta method with an integration step of $\Delta t = 0.001$. The $x$ coordinate was recorded with sampling times $\delta t \in [0.02, 1]$ with step 0.02 time units/sample. The TS length $M$ is the same for all $\delta t$. Figure 3 shows the normalized number of unobserved ordinal patterns $\mathcal{N}$ found for the entire TS as a function of the sampling time for $D = 6$ and $\tau = 1$. For the noise-free case, $\mathcal{N}$ decreases with $\delta t$, indicating the loss of temporal information due to the undersampling. In the presence of noise, the number of unobserved patterns reaches a maximum at an intermediate sampling time, $\delta t_{\text{max}} = 0.14$ time units/sample, independently of the noise level. At small sampling times the stochastic component dominates the dynamics resulting in zero unobserved ordinal patterns. For an intermediate value of $\delta t$, the nonlinear structures related to the intrinsic chaotic dynamics are detected by a maximum number of unobserved patterns. Finally, for larger sampling times, the original noisy chaotic dynamics is undersampled yielding a progressive loss of the nonlinear determinism with a decreasing $\mathcal{N}$, and the data appear to be stochastic rather than chaotic. This behavior is observed for $NL \leq 0.3$, while for a higher noise contamination the numbers of unobserved ordinal patterns are zero for all sampling times.

We note that ordinal patterns are not systems invariant like Lyapunov exponents. Their evolution with the sampling time is expected to be different according to the specific TS considered. However, TS generated by the different coordinates of a
dynamical system are expected to be characterized by ordinal patterns that do not exhibit strong discrepancies with each other [24]. Hence, the $x$ coordinate of the Lorenz equations is considered as a representative TS of the whole system. Qualitative similar results are found for TS obtained from $y$ and $z$ coordinates.

Figure 4 shows the decay of the normalized averaged $\langle N \rangle$ as a function of $w$ for two noise levels. We observe that the decay of $\langle N \rangle$ strongly depends on the sampling time. On the one hand, for $\delta t < \delta t_{N_{\text{max}}}$, the larger the sampling time the slower the decay of the unobserved ordinal patterns. On the other hand, for $\delta t > \delta t_{N_{\text{max}}}$, as the $\delta t$ grows, $\langle N \rangle$ decays faster and faster. For $\delta t = \delta t_{N_{\text{max}}}$, the slowest decay of the number of unobserved ordinal patterns with $w$ is observed. This behavior is independent of the noise level. In all cases, the decay of $\langle N \rangle$ can be appropriately fitted by a stretched exponential as can be seen in Fig. 5—solid black lines. The stretching exponent $\gamma$ and the characteristic decay rate $B$ as a function of the sampling time are depicted in Fig. 6. The $\gamma$ value reaches a minimum at $\delta t_{N_{\text{max}}}$, while the parameter $B$ shows a maximum.

IV. EXPERIMENTAL RESULTS

We now proceed to validate our numerical findings for real-world data coming from an optoelectronic oscillator with delay [19]. The experimental implementation, which includes standard telecommunication components, is depicted in Fig. 7. The optoelectronic oscillator is composed of a semiconductor laser diode feeding a Mach-Zehnder (MZ) modulator that performs a sine squared nonlinear transformation, an optical delay line (4.2 km), and an optoelectronic feedback. This feedback accounts for optical intensity detection, linear filtering, and amplification. By closing the delay loop (delay time $\sim 21$ $\mu$s), the optoelectronic feedback serves as the drive of the MZ modulator. The feedback loop has a low-pass filter with a frequency cutoff of $\sim 663$ KHz, i.e., the response time of the system is $\sim 240$ ns.
The optoelectronic oscillator with delay can generate optical chaos by following a similar approach to the pioneering work of Ikeda, as described in detail in Refs. [25,26]. This experimental system can generate intensity pulsations that are typical for the Ikeda scenario, including a period doubling route to chaos. Here, we focus on the chaotic dynamics of the system for the TS analysis. In Fig. 8, we show an example of the TS in the chaotic regime. The TS have been acquired with a sampling rate of \( f_s = 500 \) MSamples/s and a resolution of 8 bits. The measured data is long enough to obtain TS of length \( 4 \times 10^4 \) for 200 different sampling frequencies.

The experimental system depicted in Fig. 7 can be modeled with a delay differential equation. The deterministic version of the optoelectronic oscillator with delay feedback is described by [27]:

\[
\frac{dx(t')}{dt'} = -x(t') + \beta \sin^2[x(t' + \tau) + \Phi],
\]

where \( t' \) is the time in normalized units \( [t' = t/(240 \text{ ns})] \), \( \beta \) is the feedback strength, \( \tau \) is the normalized delay time, and \( \Phi \) is the phase. The parameters have been chosen to match the experimental conditions \( (\beta = 3.6, T = 87.2, \Phi = 0.85 \text{ rad}) \).

In addition to having access to the experimental TS, one advantage of further analyzing numerically generated TS is that it allows us to treat independently the deterministic and stochastic contributions. Thus, we added a stochastic component by simply adding a white Gaussian noise to the numerical TS. This additive noise has a zero mean and a standard deviation of 0.03 times the standard deviation of the original TS, yielding a signal to noise ratio compatible with the experimental data [19].

Figure 9 illustrates the normalized number of unobserved ordinal patterns as a function of the sampling time for \( D = 6 \) and \( \tau = 1 \) for the whole data set. We observe a maximum at \( \delta t_{N_{\text{max}}} \approx 80 \text{ ns} \) for both experimental data and simulation. This result evidences a noisy chaotic behavior as previously concluded by Soriano et al. [19]. The decay of the number of unobserved ordinal patterns with \( w \) is depicted in Fig. 10 for different sampling times for the experimental data. Qualitative similar behavior is observed for the numerically generated TS.

As expected, the decay of \( \langle N \rangle \) depends on the value of \( \delta t \) and the slowest decaying is observed at a sampling time near \( \delta t_{N_{\text{max}}} \).

The properties of the system are better characterized by the estimation of the fitting parameters, which are illustrated in Fig. 11. The minimum reached by the stretching exponent is around \( \delta t_{N_{\text{max}}} \). In order to interpret the meaning of \( \delta t_{N_{\text{max}}} \), we show in Fig. 12 the power spectrum of the experimental data and its numerical counterpart. It can be seen that the
FIG. 9. Normalized number of unobserved ordinal patterns $N$ as a function of the sampling time $\delta t$, using $D = 6$ and $\tau = 1$ for the optoelectronic oscillator. Experiment (open blue circles) and simulation (open red squares). The maximum of these curves is located at $\delta t_{N_{\text{max}}} = 80$ (ns). Mean and standard deviation (displayed as error bars) over ten realizations are depicted.

The decay of the chaotic dynamics merges with the noise floor at $\sim 6$ MHz. By applying the Nyquist-Shannon criterion, one can argue that a sampling frequency of approximately $f_{NS} = 12$ MHz is required to capture the complete evolution of the system. Therefore, the $\delta t$ for the slowest decay of the number of unobserved ordinal patterns, i.e., the minimum stretching exponent, is related to the sampling time needed to capture all the relevant dynamical information in the TS. The identification of such an optimum sampling time for the chaotic dynamics is also found, but less clear, when the characteristic decay rate is analyzed [Figs. 11(c) and 11(d)], reaching its maximum for $\delta t \sim \delta t_{N_{\text{max}}}$.

A good agreement is observed between the parameters estimated in the experimental [Figs. 11(a) and 11(c)] and numerical [Figs. 11(b) and 11(d)] analysis. We only find a discrepancy for small sampling times, where the stretching exponent reaches values near to one for the simulations. We rely again on the observation of the power spectrum in order to explain this discrepancy. As it can be seen in Fig. 12, the experimental data presents some high-frequency peaks, with the widest one located at $\sim 60$ MHz. We conjecture that these spurious high-frequency spectral contributions, that originate from the detection apparatus and the operational amplifiers in the feedback loop, are the reason of the observed discrepancy at small sampling times.

V. DISCUSSION

It was recently shown that the decay of the number of missing ordinal patterns follows a stretched exponential model for long-term correlated TS [10]. In the present analysis, it has
been numerically demonstrated that the same behavior can be concluded for the decay of unobserved ordinal patterns in noisy chaotic TS. In order to be able to distinguish between noisy chaotic and correlated stochastic dynamics by using the decay of the number of unobserved ordinal patterns with the sequence length, we propose a characterization by plotting the characteristic decay rate versus the stretching exponent estimated from the fitting. With this aim in mind, we have studied numerical simulations of correlated stochastic processes. Particularly, we have considered the fractional Brownian motion (fBm) and their increments, fractional Gaussian noise (fGn) [28], because they are a paradigmatic model for correlated stochastic processes.

Figure 13(a) shows the evolution of the fitting parameters as the noise level increases in the logistic map (purple diamonds) and the $x$ coordinate of the Lorenz equations (orange squares). For the Lorenz case, the parameters were selected at the optimum sampling time $\delta t_{N_{\text{max}}}$, and the stretching exponent increases with the noise level and the characteristic decay rate decreases, approaching to the stochastic zone, where the fBm processes and fGn locate. Up to NL $\sim$0.5, it is possible to distinguish between these noisy chaotic and stochastic dynamics. The values obtained at $\delta t_{N_{\text{max}}}$, for both the recorded and simulated TS of the optoelectronic oscillator are also depicted (green pentagram and cyan asterisk, respectively). Even when the noise level is maximum, NL $= 1$, the position of the quantifiers for chaotic TS in this representation plane never reaches those associated with a totally uncorrelated dynamics ($\gamma = 1$).

Figure 13(b) shows the evolution of continuous systems with the sampling time for NL $= 0.1$. We observe that the Lorenz $x$ coordinate follows a loop (clockwise) describing the interplay between the stochastic and deterministic components in the TS as the sampling time increases. A qualitatively similar behavior is observed for the experimental and simulated TS from the optoelectronic oscillator.

**VI. CONCLUSIONS**

We have proposed a mathematical model to fit the decay of the number of unobserved ordinal patterns as a function of the time series length for noisy chaotic dynamics, namely the stretched exponential decay. By numerical analysis using the logistic map and the $x$ coordinate of the Lorenz system operating in a totally developed chaotic dynamics, we showed that the model works independently of the noise contamination, and the sampling frequencies for the continuous case. Additionally, the values estimated for the parameters of the model can be used to identify the time scales for which the deterministic or stochastic nature govern the dynamics of the system. More precisely, both parameters reach the extrema at the optimum sampling time needed to capture all the nonlinear structure related to the intrinsic chaotic dynamics. Our results allow us to conjecture the discrimination between stochastic and noisy chaotic time series with up to medium noise levels (NL $= 0.5$), is possible. The distinction between highly contaminated noisy chaotic dynamics and stochastic correlated dynamics is, however, not guaranteed. Experimental application to the intensity pulsations of an optoelectronic oscillator operating in a hyperchaotic dynamics illustrates the applicability of the proposed approach in a practical setting.

We believe that our approach gives a more solid characterization and complements the original proposal by Amigó et al. [4]. Since in real-world data the chaotic dynamics will always be corrupted by observational noise, a more complete understanding of how the coexistence of determinism and noise affects the decay of the number of unobserved ordinal patterns as the sampling time increases is required.
patterns is of paramount importance. A natural extension of the current work is the detailed characterization of the influence of multiplicative noise.

ACKNOWLEDGMENTS

This work was supported by Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) FONDECYT POSTDOC Grant No. 3170533. L.Z. gratefully acknowledges financial support from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. M.C.S. was supported by the Spanish Ministerio de Economía, Industria y Competitividad through a “Ramon y Cajal” Fellowship (RYC-2015-18140). D.G.P. was supported by Pontificia Universidad Católica de Valparaíso (PUCV 039.352/19).