Intermittency in predicting the behavior of stochastic systems using reservoir computing

Nikita Kulagin [1], Andrey Andreev [1], Alexey A. Koronovskii [2], Olga I. Moskalenko [2], Aleksandr P. Sergeev [1], Artem Badarin, [1] and Alexander E. Hramov [1,*]

[1] *Engineering School of Information Technologies, Telecommunications and Control Systems, Ural Federal University, 19 Mira Street, Ekaterinburg 620002, Russia*

[2] *Physics Institute, Saratov State University, 112 Astrakhanskaya Street, Saratov 410012, Russia*

(Received 20 November 2024; accepted 23 January 2025; published 21 February 2025)

A new behavior type of reservoir computing model for predicting dynamics of stochastic systems has been observed. It has been shown that when the control parameters of the predicted stochastic system and the reservoir computing model are turned, we observe intermittent behavior, i.e., close to the threshold parameter value the reservoir computing model demonstrates the accurate prediction most of the time, but there are time intervals during which the accurate prediction is interrupted by intervals characterized by the lack of prediction. The characteristics of the intermittency in predicting the behavior of the stochastic system correspond to the well-known on–off intermittency. The concept of the effective noise to describe the quality of prediction is proposed, and the technique of its amplitude value estimation is developed.

DOI: 10.1103/PhysRevE.111.024209

I. INTRODUCTION

Intermittency in dynamical and stochastic systems is one of the most fundamental universal nonlinear phenomena in the qualitative theory of dynamical systems [1]. The phenomenon of intermittency was first described in the context of the transition to chaotic regimes in dynamical systems, where the concept of intermittency of types I–III was introduced depending on the conditions and characteristics of such a transition [2]. Noise effects in stochastic systems can also lead to the phenomenon of intermittency. Despite the randomness of such noise effects, the corresponding patterns of intermittent behavior obey general regularities. A typical example is the on–off intermittency [3,4], which is observed near a supercritical bifurcation under noise influence and is characterized by the same patterns in a wide variety of systems. The regime of on–off intermittency occurs at self-organized bistability in networks of oscillators with hierarchical organization of links (scale-free networks) [5], in human motion [6], and in biological systems associated with neuronal ensembles of the epileptic brain [7], etc.

Later it was found that the phenomenon of intermittency takes place at the boundaries of transitions to synchronization in coupled chaotic oscillators, where it shows a significant variability of the features of the transition through intermittency [8,9]. All this allowed us to introduce into consideration a whole set of intermittencies, each of which describes the behavior of coupled chaotic oscillators near the onset of the corresponding chaotic synchronization regime. Thus, at the boundary of phase synchronization, different types of intermittency are observed, depending on parameters and the topology of attractors of coupled chaotic systems, i.e., type-I intermittency, eyelet intermittency [10] (which may be considered as type-I intermittency with noise [11]), and ring intermittency [12]; whereas at the boundary of lag-synchronization and generalized synchronization, the on–off [13,14] or jump [15] intermittencies are realized, etc.

Recently, the application of machine learning techniques based on recurrent neural networks and such a variety of them as reservoir computing (RC) has attracted special interest [16,17]. RC has emerged as a powerful tool for both predicting and classifying the behavior of dynamical systems [18,19] and multivariate time series [20,21]. Its advantages include low training costs, simple architecture, and the use of fixed reservoirs. This makes reservoir computing particularly valuable for forecasting a wide array of dynamic characteristics of complex systems. Such predictive models based on RCs have proven themselves both for predicting complex dynamics of chaotic oscillators [22,23], networks [24], and spatiotemporal chaos in models described by partial derivatives [25,26]. RC offers a promising approach for predicting and analyzing the dynamic characteristics of nonlinear systems. For instance, RC can be used to calculate Lyapunov exponents [27], estimate basins of attraction [28], and predict cluster synchronization [29].

In Ref. [30] interesting assumptions were made that the RC is effective for predicting the temporal dynamics when it is in a regime of generalized synchronization with the original system. Therefore, we can expect that near the regime of accurate prediction of the behavior of the system under study it is possible to expect an intermittency effect to be observed. Moreover, taking into account the information about the behavior of systems near the boundary of generalized synchronization (or noise-induced synchronization as a special case of generalized synchronization [31]), we can assume that such behavior should obey the law of on–off intermittency.

*Contact author: hramovae@gmail.com

The purpose of the presented work was to verify this assumption. We consider one of the RC-based models proposed in Ref. [32] to predict the behavior of a stochastic FitzHugh-Nagumo neuron over a wide range of driven noise parameters. We show that an intermittency effect, which has some similarities with on–off intermittency, is observed in predicting the behavior of a stochastic neuron when exposed to the same noise signal that affects the original FitzHugh-Nagumo model described by stochastic differential equations (SDEs). We propose also the concept of the effective noise to describe the quality of prediction based on the technique of the effective noise amplitude value estimation.

II. METHODS

A. Stochastic FitzHugh-Nagumo neuron

To investigate the intermittency effect in the RC-based predictive model, we consider the stochastic FitzHugh-Nagumo neuron [33], which is often used in various studies of neural ensembles, as the basic system under study. This model also demonstrates the effect of stochastic resonance in neural ensembles [34], which has also been confirmed experimentally, for example, in human visual perception [35]. The model is described by the following system of SDEs:

$$\begin{aligned}\dot{\eta}_1 &= \eta_1 - \eta_1^3/3 - \eta_2 + \beta,\\ \dot{\eta}_2 &= 0.08(\eta_1 - 0.8\eta_2 + 0.7) + D_n\xi(t),\end{aligned} \tag{1}$$

where η_1 and η_2 are the excitatory and recovery variables, respectively.

The model (1) contains a stochastic term, denoted as $D_n\xi(t)$, representing zero-mean white Gaussian noise. This noise exhibits an autocorrelation function defined by $\langle\xi(t)\xi(t+\tau)\rangle = \delta(\tau)$, where D_n signifies the noise amplitude. The parameter β plays a crucial role in determining the equilibrium points of the system, directly impacting the neuron's threshold for excitation. For our study, we investigate the excitable regime, specifically focusing on the precritical state. This state undergoes a Hopf bifurcation at approximately $\beta_c = 0.322$, and below β_c the model neuron does not display self-sustained oscillations. To ensure operation of the neuron within the excitable regime, we set the parameter β to 0.3, which corresponds to a value well within the range $0 < \beta < \beta_c$.

In the absence of noise influence (i.e., $D_n = 0$), the system (1) reaches a stable stationary state. However, as the noise intensity (D_n) increases, the system begins to generate the irregular sequence of spikes. The characteristics of this sequence of spikes are dependent on the noise intensity.

For numerical solution of the SDEs (1), we used the Euler-Maruyama method [36] with a time step of integration set at $\Delta t = 0.1$.

B. RC-based model for predicting the dynamics of a stochastic neuron

In our previous work [32] we considered an RC-based model that is capable of predicting the behavior of a system described by SDEs but also allows us to predict the behavior of such a system over a wide range of control parameters. For example, when varying the noise intensity D relative to the value D_0 at which the model was trained, the RC predicts a resonant change in the degree of regularity of the spike sequence generated by the stochastic neuron. In other words, it predicts the effect of stochastic resonance, despite the fact that this information was not provided to it when the reservoir computer was trained. These points are discussed in more detail in Ref. [32].

As it is well known, RC is a kind of recurrent neural network, which has three layers in its structure, i.e., input, hidden (reservoir), and output layers, in which only the weights of the links from the hidden layer to the output one are trained. Due to this fact the RC is a simple and easily trainable machine learning model.

In our case, following Ref. [32], the input layer has three inputs, corresponding to an input vector

$$\mathbf{I}(t) = (\eta_1(t), \eta_2(t), D_r\xi(t))^T, \tag{2}$$

which includes two independent variables of the SDEs under study and a noise signal ξ_t at time t that drives the stochastic neuron with the amplitude D_r. Each of the input signals acts on n_1 unique neurons in the reservoir layer with pregenerated random weights given by the input matrix $\mathbf{V}$ with dimension $3 \times N$, where $N = 3n_1$ is the total number of artificial neurons in the reservoir layer. So, each jth reservoir's neuron receives only one input signal from only one of three input neurons with random input weight $V_{i,j}$, where i is the index of the input neuron. For example, if the jth neuron is connected to the first input one, than $V_{1,j} = v_{1,j}$, $V_{2,j} = V_{3,j} = 0$, where each $v_{i,j}$ is randomly generated from the uniform distribution $[-1, 1]$.

The state of the reservoir layer at each moment of time t is defined by the following mapping,

$$\mathbf{R}(t) = \tan h(\mathbf{QR}(t-\Delta t) + \mathbf{VI}(t)), \tag{3}$$

where $\mathbf{Q}$ is the reservoir layer's adjacency matrix, which sets the weights of connections between N artificial neurons. The matrix $\mathbf{Q}$ defines a random network characterized by the average node degree $\langle k\rangle$ and the spectral radius λ.

The output vector

$$\mathbf{G}(t+\Delta t) = [g_1(t+\Delta t), g_2(t+\Delta t)]^T \tag{4}$$

includes only two predicted values corresponding to SDEs variables (η_1, η_2) at the next moment of time $(t+\Delta t)$. The task of predicting the noise is not set, and it is impossible in principle, so the third noise variable is absent in the output layer. The values of the output vector $\mathbf{G}$ at time $t+\Delta t$ are formed from the values taken by the artificial neurons in the reservoir layer as $\mathbf{G}(t+\Delta t) = \mathbf{W}\hat{\mathbf{R}}(t)$, where the matrix $\mathbf{W}$ defines the weights of the output layer and the matrix $\hat{\mathbf{R}}(t)$ is the augmented reservoir layer state $\mathbf{R}(t)$, represented as vector with N components $\hat{r}_i(t) = r_i^2(t)$ for even i and $\hat{r}_i(t) = r_i(t)$ for odd i $(i = \overline{1, N})$ [28]. This augmentation increases the model's complexity by squaring the hidden state in half of the nodes, leading to improved performance on complex tasks. The output matrix $\mathbf{W}$ is calculated during the training process.

In the training mode, the RC operates in a nonautonomous mode. We feed a sequence $\mathbf{I}(t)$ consisting of true values of the stochastic neuron state $\eta_1(t), \eta_2(t)$ and a noise signal $D_r\xi(t)$ to the input of the RC and determine the weights of the output layer ($\mathbf{W}$) by minimizing the L_2 error between the

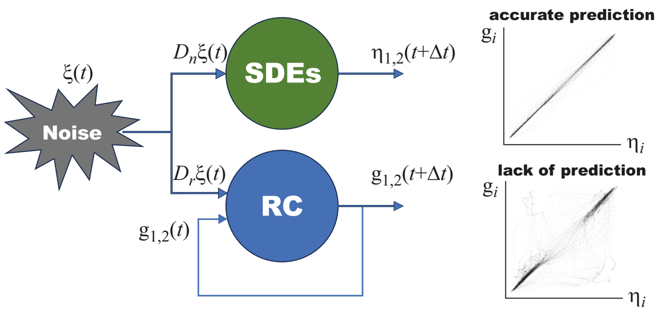

FIG. 1. Basic scheme for observing intermittent behavior in predicting stochastic dynamics using a trained RC. We drive both the stochastic system (1) and the RC (3) with the same noise signal $\xi(t)$ ($D_r = D_n$). In the case of accurate prediction, we will observe a diagonal in coordinates ($\eta_{1,2}, g_{1,2}$); in the case of lack of prediction we will observe a point cloud. In the case of noise amplitude tuning, $\Delta D \neq 0$, intermittent behavior will be characterized by switching between these modes (prediction / no prediction) in time.

true value $\mathbf{I}(t)$ and the predicted state $\mathbf{G}(t)$ using Tikhonov regularization:

$$L_2 = \sum_{t=1}^{T_{\text{train}}} ||\mathbf{G}(t) - \mathbf{I}_{1,2}(t)||^2 + \gamma ||\mathbf{W}||^2, \qquad (5)$$

where $\gamma = 10^{-4}$ is the regularization hyperparameter, and T_{train} is the duration of the training process.

In the prediction mode, the RC (determined in the training mode output layer matrix $\mathbf{W}$) operates autonomously by feeding at each discrete time step $t + \Delta t$ the RC output values to the RC input, i.e., $\eta_{1,2}(t + \Delta t) := g_{1,2}(t + \Delta t)$. At the same time, a noise signal $D_r\xi(t)$ should also be applied to the RC input as the third reservoir input.

C. Analysis of intermittency in RC-based prediction

Note, to predict the dynamics of a stochastic neuron, it is not enough just to train a RC. It is also necessary to provide information to the trained RC about the noise source that drives the original system of SDEs. The question of prediction accuracy is most simply addressed using the scheme shown in Fig. 1. We have an initial model described by stochastic differential equations (SDEs in Fig. 1) and trained RC (RC in Fig. 1). To predict the SDEs signal using RC, we need to know the actual noise influence on the SDEs system, so we generate the single noise signal $D\xi(t)$ and feed it to both models, i.e., the predicted SDEs system and the RC-based model for prediction. The prediction quality can be evaluated using the root-mean-square deviation between the true and predicted values:

$$\delta(t) = \sqrt{(\eta_1(t) - g_1(t)^2) + (\eta_2(t) - g_2(t))^2}. \qquad (6)$$

An accurate prediction will be characterized by the value $\bar{\delta}_T \approx 0$, which is $\delta(t)$ averaged over a long time interval T: $\bar{\delta}_T = (1/T)\int_t^{t+T} \delta(t)\,dt$. From the mathematical point of view, the condition of proximity to zero means the introduction of some small threshold $\delta_0 > 0$; if the value $\bar{\delta}$ does not exceed it, we can talk about an accurate prediction of the SDEs behavior in time:

$$\bar{\delta}_T < \delta_0. \qquad (7)$$

Lack of prediction means in this case the opposite condition

$$\bar{\delta}_T > \delta_0. \qquad (8)$$

However, let us hypothetically assume that we can observe a borderline situation between accurate prediction and lack of prediction: at some time intervals L_i we have an accurate prediction $\bar{\delta}_{L_i} < \delta_0$, which is then followed by time intervals S_i where the regime of lack of prediction $\bar{\delta}_{S_i} > \delta_0$ is observed. The sum of all intervals makes up the total observation time $\sum_i S_i + \sum_i L_i = T$, during which we follow the prediction. By analogy with intermittent generalized synchronization, this intermittent behavior can be expected with additional small parameter upset, e.g., we can feed the original stochastic neuron (1) and the RC by noise with slightly detuned values of amplitudes D_n and D_r, respectively, where the parameter mismatch is $\Delta D = D_r - D_n$, ($D_n \neq D_r$).

Obviously, such a regime carries the features of intermittency, where the L_i intervals of the accurate prediction correspond to the laminar phases, while the S_i intervals are the outputs from the accurate prediction regime calling the turbulent phases. Let us consider in more detail the possibility of intermittent prediction of the stochastic system behavior in the next section of the paper.

III. RESULTS

The construction of an RC model for predicting dynamics always starts with the selection of the hyperparameters of the reservoir layer network given by the matrix $\mathbf{Q}$. We have chosen the size of the reservoir layer equal to $N = 501$ artificial neurons. For training, we used the noise signal with amplitude $D_0 = 0.2$ and the signals of the stochastic neuron, which were obtained at the same value of the noise amplitude fed to the neuron, i.e., $D_n = D_0 = 0.2$. We considered the two-dimensional hyperparameter space ($\langle k\rangle, \lambda$), where the following parameter ranges were considered: average node degree $10 \leqslant \langle k\rangle \leqslant 20$ and spectral radius $0.1 \leqslant \lambda \leqslant 1.9$. It should be noted that the chosen value of the noise amplitude D_0 corresponds to the coherent resonance mode, when the signal of the neuron in the form of a sequence of generated spikes is characterized by a pronounced regularity in noise-induced oscillations [34,37].

Optimization of the hyperparameters was performed by a grid search and accordingly trained 209 RCs, among which we selected the best one in terms of the accuracy of prediction. Figure 2 illustrates the distribution of prediction quality $\bar{\delta}_T$ (6) of all trained RCs with different pairs ($\langle k\rangle, \lambda$) of hyperparameters. It is clearly seen that there is the best RC which exhibits the highest prediction quality (indicated by the arrow in Fig. 2). The reservoir layer hyperparameters for this RC are the average node degree $\langle k\rangle = 20$ and the spectral radius $\lambda = 1.9$. It is this RC that we will further use to analyze the intermittency effect in predicting the behavior of a stochastic neuron.

Let us now consider the characteristics of the predicted stochastic process when the parameters of the external noise influence are disordered between the original system and the

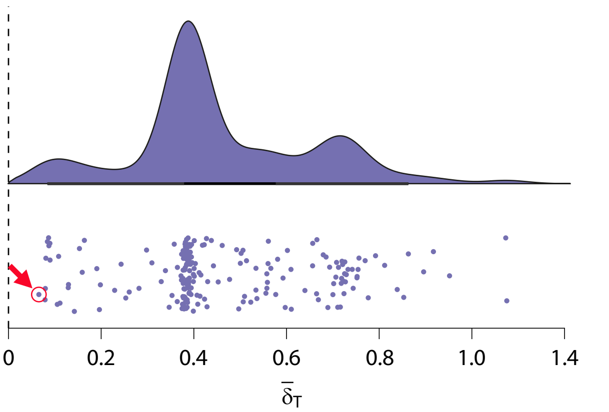

FIG. 2. The distribution of prediction quality $\bar{\delta}_T$ (6) of all RCs using for hyperparameters selection. The RC which exhibits the highest prediction quality is indicated by the red arrow.

trained RC. Let us choose the value of the noise amplitude fed to the neuron $D_n = 0.2$. Recall that the RC was trained at the same value of the noise intensity $D_0 = 0.2$. We will adjust the intensity of the noise D_r fed to the RC from a fixed value D_n and investigate the quality of the prediction. Figure 3(a) shows the dependencies of the predicted signal $g_1(t)$ and the original signal $\eta_1(t)$, as well as the difference of these signals $g_1(t) - \eta_1(t)$ over time for noise amplitude $D_r = 0.32$. The analysis of characteristics shows that there is a behavior similar to intermittency, when we observe laminar phases L_i, during which we have an accurate prediction, and phases S_i, when the time series diverges and an accurate prediction on these time intervals is not observed (e.g., for S_i phases the generation of "useless" spikes takes place). It is interesting to note that the amplitude of irregular bursts on the difference $g_1(t) - \eta_1(t)$ is large enough and exceeds the amplitude of spikes generated by the stochastic neuron.

To quantify the statistical properties of intermittent behavior, we compute the distribution $N(L)$ of the laminar phase durations L. We suppose that the current time interval L is the laminar phase if the condition $\delta(t) < \delta_0$ is satisfied $\forall t \in L$ (where $\delta_0 = 0.1$); otherwise the current state is assumed to be an turbulent phase. Note, the criterion for laminar and turbulent phase detection used in the numerical simulation is somewhat more stringent in comparison with (7) and (8), which allows us to guarantee the high quality of prediction at limited time intervals L. Figure 3(b) shows the distribution of the laminar phase durations plotted on a log–log scale. One can see that this distribution is close to the power law with exponent $\alpha = -1.5$. It should be noticed that this result does not sensitively depend on the value of threshold δ_0. This form of the distribution of laminar phase durations indicates the presence of on–off intermittency.

The other criterion of the on–off intermittency is the dependence of the mean laminar phase duration $\langle L\rangle$ on the deviation of the control parameter from the critical value. In our case, the value of the noise amplitude at which the reservoir was trained, $D_0 = 0.2$, was chosen as the critical value of the control parameter. At the same time, one must realize that even for the best RC we do not get a perfect prediction where $\delta(t) = 0$. Therefore, even for the best RC we have a situation that sometimes small failures in prediction can occur, which we will further take into account when analyzing the dependencies of the mean laminar phase duration on the noise parameters. It should be also noted that this is a fundamental limitation, since it is impossible to create a finite-size reservoir that would perfectly $[\delta(t) = 0]$ describe the behavior of the original SDEs.

Let us first consider the situation when we tune the noise parameter of a stochastic neuron, D_n, whereas the reservoir predicts the neuron's behavior at the value $D_r = D_0 = 0.2$ at which it was trained, and hence the prediction is the most accurate. In this case, the deviation is represented as $\Delta D = (D_n - D_0)$. We vary the noise amplitude D_n in the range from 0.2 to 0.4. Figure 4(a) shows the determined mean laminar phase length $\langle L\rangle$ versus deviation ΔD on a log–log scale. From this figure one can see the universal power law $\langle L\rangle \sim [\Delta D]^\rho$ with critical exponent ρ being very close to -1. This is further evidence that the observed intermittency in predicting the behavior of a stochastic system is on–off intermittency.

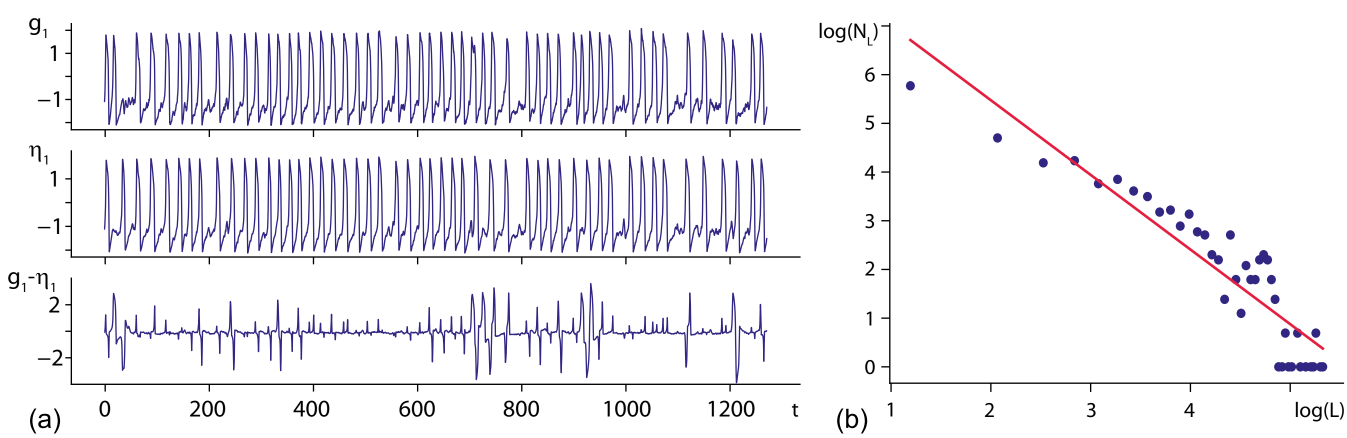

FIG. 3. (a) The dependencies of the predicted signal $g_1(t)$, the original signal $\eta_1(t)$, and the difference of these signals $g_1(t) - \eta_1(t)$ over time. (b) The statistical distribution of laminar phases L and its approximation $N(L) \sim L^\alpha$, where $\alpha = -1.53$, in log–log scale. This form of the distribution of laminar phase durations indicates the presence of on–off intermittency. The noise amplitudes are $D_0 = 0.2$, $D_n = 0.2$, $D_r = 0.32$, the value of the threshold $\delta_0 = 0.1$.

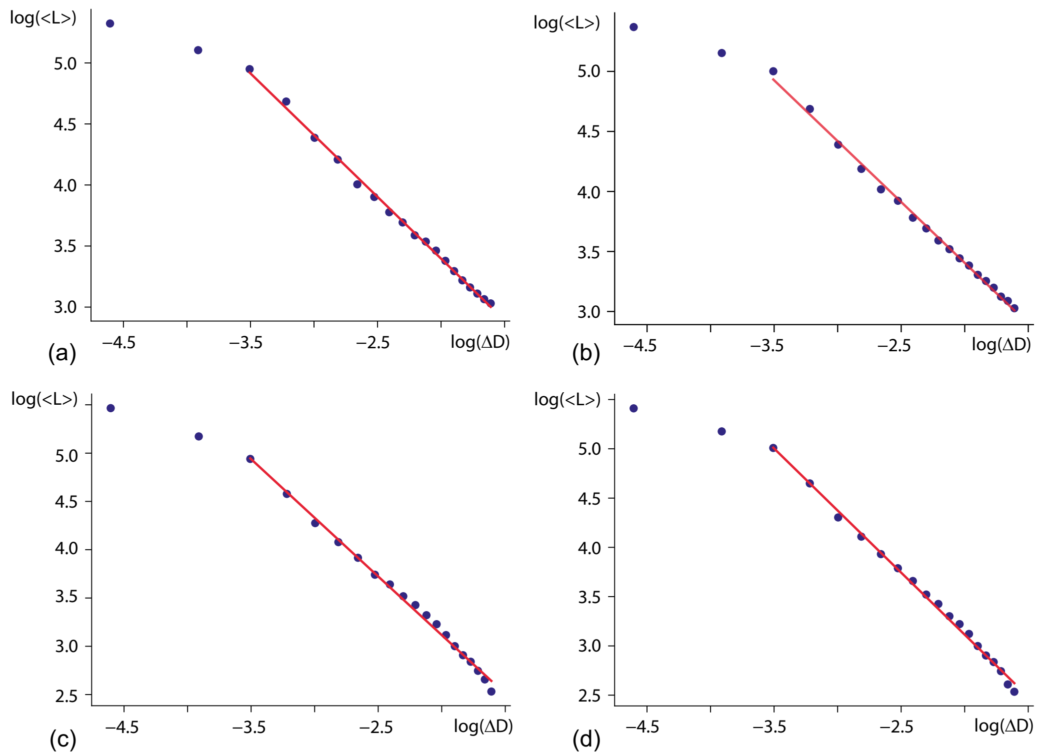

FIG. 4. The log–log plots of the mean laminar phase duration $\langle L\rangle$ vs the deviation ΔD (points) and their approximations by power law $\langle L\rangle \sim [\Delta D]^{\rho}$ (solid red line) for following specific situations. (a) The RC noise parameter is fixed $D_r = D_0 = 0.2$, the neuron noise parameter D_n is varied. The critical exponent is $\rho = -1.01$. (b) The RC is replaced by an auxiliary neuron. The critical exponent is $\rho = -1.01$. (c) The RC noise parameter D_r is varied, the neuron noise parameter is fixed $D_n = D_0$. The critical exponent is $\rho = -1.21$. (d) The original neuron is replaced by an auxiliary RC. The critical exponent is $\rho = -1.26$. For all situations, we observe linear dependencies in log–log scale, but only (a) and (b) conform to the on–off intermittency law when we change the noise parameter of the neuron.

We can check our result on a modeling situation. In fact, the RC is a digital twin [38] of the stochastic neuron, so we can consider a situation similar to the previous one by replacing the RC with an auxiliary neuron [being identical to (1)] to which we will apply the same noise signal $\xi(t)$ with the amplitude D_r. In fact, in this case we consider the effect of intermittency between two neurons (1) that have the same noise source $\xi(t)$ but different noise amplitudes: $D_r = D_0 = 0.2$ and $D_n \in [0.2, 0.4]$, respectively. Figure 4(b) illustrates the mean laminar phase length $\langle L\rangle$ versus deviation ΔD for this case. The universal power law $\langle L\rangle \sim [\Delta D]^{-1}$ is again observed. We see that the obtained results are in full agreement with previous results, i.e., the RC accurately models the behavior of a neuron when the noise parameter is tuned and predicts on–off intermittency between two neurons.

It should be noted that we can also vary the control parameter, not only of the neuron but also of the RC. In this case we will vary the noise amplitude D_r, which is fed to the RC, and the deviation parameter has the form $\Delta D = D_r - D_0$, whereas the intensity of noise being fed to the neuron (1) remains fixed as $D_n = D_0 = 0.2$. The obtained dependence of the laminar phase duration on the supercriticality parameter ΔD for this case is shown in Fig. 4(c). It is well seen that we also observe a power law, however, with the exponent $\rho = -1.21$. Interestingly, if we replace the original stochastic neuron with an auxiliary RC (being the better RC among all RCs), to which we apply a noise signal with amplitude D_0 at which the RC was trained, then as can be seen from Fig. 4(d), we again have a power law but with exponent $\rho = -1.26$. The slight difference between the results shown in Figs. 4(c) and 4(d) is probably due to the fact that replacing a neuron with an RC leads to additional rare prediction inaccuracies, which we mentioned above when discussing the nonideal prediction of even the best RC among all trained ones at noise amplitude D_0. Thus, changing the RC parameters relative to those at which the RC was trained leads to a deviation from the "ideal" power law for on–off intermittency with -1 exponent when modeling two neurons directly [see Fig. 4(b)].

We assume that this deviation of exponent ρ from the "classical" value of "minus one" is caused by an error in predicting the neuron dynamics by the RC-based model. Indeed, as we discussed above, an RC trained at noise intensity D_0 cannot perfectly predict the neuron dynamics for $D_r \neq D_0$. In other words, if noise signals with identical amplitudes $D_n = D_r > D_0$ are applied to the SDEs and the RC, respectively, the predicted dynamics contain more errors compared to the case of $D_n = D_r = D_0$. Obviously, one can expect that the more the noise intensity D_r (and, correspondingly, D_n too) differs from the base amplitude D_0, the worse RC predicts the dynamics of the neuron. To illustrate this aspect, Fig. 5 shows how the averaged laminar phase duration $\langle L\rangle$ depends on the RC noise intensity D_r for the case when $D_r = D_n$. From Fig. 5, which is plotted in the semilogarithmic scale, we see that the curve $\langle L\rangle(D_r)$ has an decreasing exponential

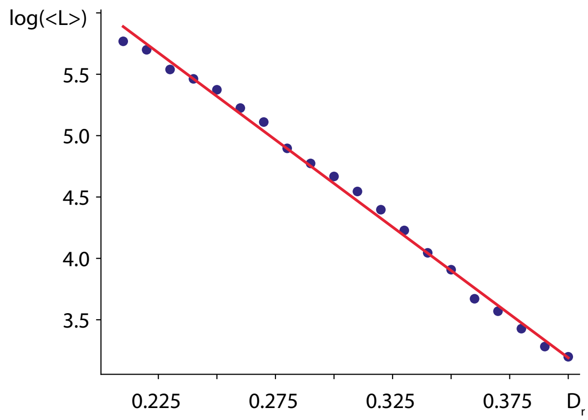

FIG. 5. The semilog plot of the mean laminar phase duration $\langle L\rangle$ vs the RC noise intensity D_r (points) and their approximations by exponential law $\langle L\rangle \sim e^{-kD_r}$ (solid red line), $k = 13.68$. Here $\langle L\rangle$ is estimated for each $D_r = D_n$. Thus, the more the noise intensity D_r (and thus D_n, too) differs from the baseline amplitude D_0, the worse RC predicts the dynamics of the stochastic neuron.

character with an exponent $k = 13.68$. As can be seen, the larger $(D_r - D_0)$ is, the smaller the average laminar phase duration $\langle L\rangle$ is, with the average laminar phase duration decreasing exponentially.

So, we suppose that in Figs. 4(c) and 4(d), changing D_r leads to decreasing $\langle L\rangle$ because of two reasons: (i) increasing the difference between RCs, D_r, and neurons, D_n, noise amplitudes, and (ii) increasing self-errors of the RC. Then the difference between an "ideal" exponent of "minus one" and a real exponent of power laws in Figs. 4(c) and 4(d) may be caused by the second mechanism, and, therefore, we have to estimate it quantitatively and consider this point in more detail.

IV. ANALYTICAL ESTIMATIONS OF RC SELF-ERRORS

To estimate RC self-errors, both analytically and numerically, let us start with several assumptions and refinements:

(1) We assume that the inaccuracies in prediction of RC may be characterized both qualitatively and quantitatively with the help of the additional effective noise $\eta(t)$ with the same characteristics as in stochastic FitzHugh-Nagumo neuron model (1) (i.e., zero-mean white Gaussian noise with an autocorrelation function defined by $\langle\eta(t)\eta(t+\tau)\rangle = \delta(\tau)$) and amplitude D^*.

(2) To distinguish the cases considered in Sec. III and shown in Figs. 4(a) and 4(b), and Figs. 4(c) and 4(d), we refine the notations for the noise amplitude differences as $\Delta D_n = D_n - D_0$ and $\Delta D_r = D_r - D_0$, respectively.

(3) We assume that the amplitude of the effective noise D^* depends only on the difference between the external stochastic signal amplitude D_r fed to the RC and the noise intensity D_0 used to train the RC. Additionally, we also assume that D^* does not depend on the base amplitude D_0. This is a rather rough assumption, but the results of our numerical simulations indicate that for the purposes of current consideration such an approximation can be used. In other words, we operate in the paradigm that $D^* = D^*(\Delta D_r)$.

(4) $D_0^* = D^*(0)$ is assumed to be very small (i.e., $D_0^* \ll D_0$), and with the growth of ΔD_r the value of D^* is supposed to increase.

(5) We also refine the difference ΔD between the amplitudes of signals fed to SDEs and the RC as

$$\Delta D^* = |D_n - D_r| + D^* \tag{9}$$

to take into account the introduced effective noise describing prediction self-errors of RC caused by differences of the stochastic signal amplitudes used in the current calculation (D_r) and training (D_0).

(6) All stochastic signals [i.e., $D_n\xi(t)$, $D_r\xi(t)$, $D^*\eta(t)$] are considered as equal in terms of their influence on the prediction quality and duration of laminar phases.

Taking into account the assumptions and refinements made above, one can write the power law for the laminar phase distribution given in Fig. 4(a) as

$$\langle L\rangle = A(D_n + D_0^* - D_0)^\rho = A(\Delta D_n + D_0^*)^\rho. \tag{10}$$

Since $D_0^* \ll D_0 < D_n$, the insufficient fluctuations of parameters A and ρ (as well mean laminar lengths $\langle L\rangle$) in comparison with case shown Fig. 4(a) may be neglected, and their values may be obtained from the numerical data presented in Fig. 4(a).

The exponential law given in Fig. 5, in turn, takes the form

$$\langle L\rangle = B\exp(-k\Delta D_r). \tag{11}$$

Note, Fig. 4(a) is obtained for $\Delta D_n > 0$ and $\Delta D_r = 0$, which is why D_0^* is substituted in Eq. (10), whereas for Fig. 5 $\Delta D_r > 0$ is used. Accordingly, relations (10) and (11) turn out to be simultaneously valid only for $\Delta D_n = \Delta D_r = 0$. In this case (e.g., for $\Delta D_n = \Delta D_r = 0$), from Eqs. (10) and (11) one can obtain

$$D_0^* = \exp\left(\frac{\log B - \log A}{\rho}\right). \tag{12}$$

The approximations of numerical data in Figs. 4(a) and 5 give $\log A \approx 1.37$, $\log B \approx 8.87$, and $\rho \approx -1.01$. Therefore, the amplitude of the effective noise may be estimated as $D_0^* \approx 6\times 10^{-3}$. The found value of D_0^* corresponds to the case when noise with the amplitude D_0 (i.e., with the amplitude at which the model was trained) fed both the neuron and the RC.

Let us make a few more assumptions:

(7) Due to the observed regularities inherent in intermittency of the on–off type and the closeness of the parameter ρ obtained by approximation (see Fig. 4(a) to the value of –1, we will further use $\rho = -1$.

(8) Equation (10) may be generalized in the form

$$\langle L\rangle = A(\Delta D^*)^{-1} \tag{13}$$

and, due to Assumption 6, Eq. (13) is believed to be correct in all considered cases.

Based on Assumption 8, we can apply Eq. (13) to Fig. 5. As a consequence, Eqs. (9), (11), and (13) give

$$D^*(\Delta D_r) = D_0^* \exp(+k\Delta D_r). \tag{14}$$

The obtained results may be verified numerically with the help of data sets used in Figs. 4(a) and 5. Indeed, if law (13)

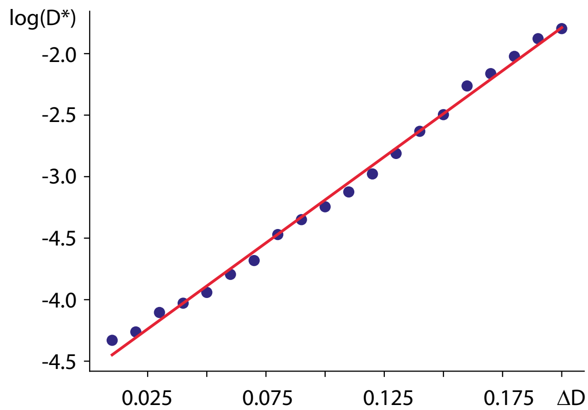

FIG. 6. The semilog plot of the effective noise intensity D^* vs the deviation ΔD_r (points) and their approximations by exponential law (14) (solid red line), $k = 13.69$. Increasing the deviation ΔD_r leads to an exponential increase of the effective noise intensity D^*.

is universal, the dependencies in Figs. 4(a) and 5 must be the same but are plotted in different coordinates of the abscissa axis ($\log \Delta D$, which in the considered case is practically equivalent to $\log \Delta D^*$ for Fig. 4(a), see also explanations to Eqs. (10) and (11) given above, and D_r, which may be easily transformed to ΔD_r in Fig. 5).

Then one can use values of the mean laminar phase duration $\langle L\rangle$ from Fig. 5 and obtain the values of the effective noise amplitude D^* corresponding to each D_r. The obtained dependence is illustrated in Fig. 6. As one can see, such dependence has an exponential form (linear in the semilog scale) that agrees well with Eq. (14). Increasing the deviation ΔD_r leads to exponential increase of the effective noise intensity D^*.

Then, we rescale the deviation by taking into account the calculated effective noise intensity as Eq. (9). So, we can replot the dependencies from Figs. 4(c) and 4(d) using the corrected deviation ΔD^*. The corresponding dependencies are shown in Fig. 7: part "a" is the rescaled Fig. 4(c) when the neuron noise intensity is fixed and we vary the RC noise amplitude; part "b" is the rescaled Fig. 4(d) when $\langle L\rangle$ is calculated for two RCs with different noise intensities. As one can see from these figures, there are linear dependencies in log–log scale with critical exponent close to "-1" instead of "-1.21" and "-1.26" observed for common unrescaled deviation [compare with Figs. 4(c) and 4(d)]. So, it proves our suggestion about the influence of RC's self-errors on the mean laminar phase duration and confirms that in a wide range of control parameters of noise we observe the on–off intermittency effect when predicting the dynamics of the model described by the stochastic differential equations.

It should be also noted that Figs. 4(a) and 4(b), and 7(a) and 7(b) are practically identical, which confirms the universality of power law (13).

V. CONCLUSIONS

We have observed the new behavior type of an RC-based model for predicting the dynamics of stochastic systems. This mode consists in predicting the behavior of the excited

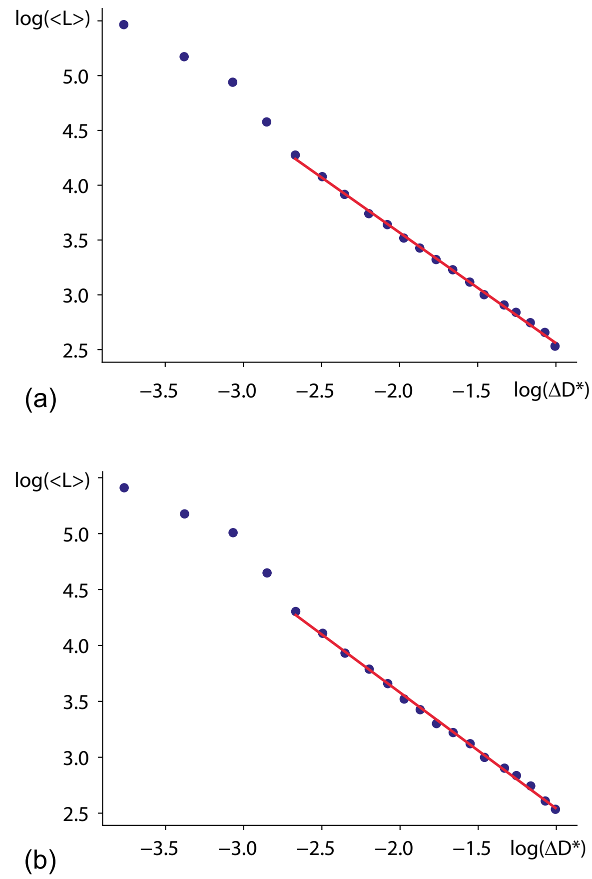

FIG. 7. The log–log plots of the mean laminar phase duration $\langle L\rangle$ vs the corrected deviation ΔD^* (points) and their approximations by power law $\langle L\rangle \sim [\Delta D^*]^{\rho}$ (solid red line) for the following specific situations: (a) Rescaled Fig. 4(c), the RC noise parameter D_r is varied, the neuron noise parameter is fixed $D_n = D_0$. The critical exponent is $\rho = -1.01$. (b) Rescaled Fig. 4(d), the original neuron is replaced by an auxiliary RC. The critical exponent is $\rho = -1.04$. Using the corrected deviation ΔD^* allows us to refine the magnitude of the supercriticality parameter and observe the effect of on–off intermittency for all considered situations.

stochastic system, not for the whole prediction time but only for some time intervals, which are randomly switched to the mode of no prediction and vice versa. A certain connection of this type of intermittent prediction with the regimes of intermittent generalized synchronization in unidirectionally coupled chaotic oscillators is revealed [14]. This is confirmed by the fact that the characteristics of the intermittent behavior correspond to the on–off type of intermittency, which is also observed at the boundary of generalized synchronization and noise-induced synchronization.

The findings of this study possess significant practical implications for the design of stochastic system models using reservoir computing, as suggested in Ref. [32]. We believe that intermittent prediction, as observed here, may be a universal phenomenon applicable to a broad range of stochastic systems analyzed with recurrent neural network technologies. Furthermore, we propose that this approach is not limited to the Gaussian noise investigated in this work but extends to other noise types, including colored, Lévy, and multiplicative noise, given the established observation of on–off intermittency across diverse noise characteristics [39,40].

The phenomenon of on–off intermittency observed in coupled systems offers valuable insight into the underlying mechanisms of extreme event formation. Specifically, the paper of Pyragas and Pyragas [41] demonstrated the efficacy of RC in predicting extreme events within both globally coupled FitzHugh-Nagumo neurons and a system of two nearly identical, unidirectionally coupled chaotic oscillators. These findings suggest that RC-based models possess the capability to predict extreme events induced by external noise signals.

The concept of the effective noise to describe the quality of prediction is proposed and the technique of its amplitude value D^* estimation is developed. Probably, this quantity may be used to characterize the quality of prediction. Perhaps it may be correlated with the averaged root-mean-square deviation between the true and predicted values $\bar{\delta}_T$.

It should be also noted that the results obtained in our paper allow us to draw the following firm conclusions. First of all, the successful prediction over a long time interval for the dynamics of a nonlinear system being under an external random (or chaotic) signal is possible if and only if the system under study is in the regime of noise-induced (or generalized, respectively) synchronization. In this case, the inaccuracy of the model, characterized by effective (internal) noise, leads to the above-described intermittency where time intervals with good prediction are replaced by areas with no prediction and vice versa. Conversely, when the system under study is in the asynchronous regime (i.e., not in the noise-induced or generalized synchronization regime), the prediction of its dynamics is impossible on a long time interval due to Lyapunov instability. An intermediate variant is also possible, when the system under study itself demonstrates the regime of on–off intermittency (see, e.g., [42,43]). In this case the prediction of the dynamics of the system under study may be obtained only for laminar phases (areas of the synchronous behavior). One should also take into account the presence of the effective (internal) noise described above in our paper and the phenomenon of multistability observed recently [44,45] for the intermittent noise-induced (or generalized, respectively) synchronization near the boundary of the synchronous regime onset. Assuming the existence of a good model (in the considered case, the RC-based model), both of these factors may lead to the occurrence of alternating time intervals with the presence and absence of prediction during laminar (synchronous) phases of the behavior of the system under study.

Another important issue here is the question of intermittent behavior at the boundary of synchronous states when the control parameters of influence on each of the systems are upset. This problem is beyond the scope of our study but, nevertheless, requires its solution in the future.

ACKNOWLEDGMENT

The research funding from the Ministry of Science and Higher Education of the Russian Federation (Ural Federal University Program of Development within the Priority-2030 Program) is gratefully acknowledged.

All authors contributed equally to this work.

[1] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani, J. Phys. A: Math. Gen. **18**, 2157 (1985).

[2] H. G. Schuster and W. Just, *Deterministic Chaos: An Introduction* (John Wiley & Sons, New York, 2006).

[3] N. Platt, E. A. Spiegel, and C. Tresser, Phys. Rev. Lett. **70**, 279 (1993).

[4] J. F. Heagy, N. Platt, and S. M. Hammel, Phys. Rev. E **49**, 1140 (1994).

[5] N. Frolov and A. Hramov, Phys. Rev. E **106**, 044301 (2022).

[6] J. L. Cabrera and J. G. Milton, Phys. Rev. Lett. **89**, 158702 (2002).

[7] A. Hramov, A. A. Koronovskii, I. Midzyanovskaya, E. Sitnikova, and C. Van Rijn, Chaos **16**, 043111 (2006).

[8] E. Rosa, E. Ott, and M. H. Hess, Phys. Rev. Lett. **80**, 1642 (1998).

[9] G. V. Osipov, B. Hu, C. Zhou, M. V. Ivanchenko, and J. Kurths, Phys. Rev. Lett. **91**, 024101 (2003).

[10] A. Pikovsky, G. Osipov, M. Rosenblum, M. Zaks, and J. Kurths, Phys. Rev. Lett. **79**, 47 (1997).

[11] A. E. Hramov, A. A. Koronovskii, M. K. Kurovskaya, A. A. Ovchinnikov, and S. Boccaletti, Phys. Rev. E **76**, 026206 (2007).

[12] A. E. Hramov, A. A. Koronovskii, M. K. Kurovskaya, and S. Boccaletti, Phys. Rev. Lett. **97**, 114101 (2006).

[13] S. Boccaletti and D. L. Valladares, Phys. Rev. E **62**, 7497 (2000).

[14] A. E. Hramov and A. A. Koronovskii, Europhys. Lett. **70**, 169 (2005).

[15] A. A. Koronovskii, O. I. Moskalenko, A. A. Pivovarov, V. A. Khanadeev, A. E. Hramov, and A. N. Pisarchik, Phys. Rev. E **102**, 012205 (2020).

[16] S. Shahi, F. H. Fenton, and E. M. Cherry, Mach. Learn. Appl. **8**, 100300 (2022).

[17] M. Yan, C. Huang, P. Bienstman, P. Tino, W. Lin, and J. Sun, Nat. Commun. **15**, 2056 (2024).

[18] D. J. Gauthier, E. Bollt, A. Griffith, and W. A. Barbosa, Nat. Commun. **12**, 5564 (2021).

[19] M. U. Kobayashi, K. Nakai, Y. Saiki, and N. Tsutsumi, Phys. Rev. E **104**, 044215 (2021).

[20] F. M. Bianchi, S. Scardapane, S. Løkse, and R. Jenssen, IEEE Trans. Neural Networks Learn. Syst. **32**, 2169 (2020).

[21] J. Moon, W. Ma, J. H. Shin, F. Cai, C. Du, S. H. Lee, and W. D. Lu, Nat. Electron. **2**, 480 (2019).

[22] A. Griffith, A. Pomerance, and D. J. Gauthier, Chaos **29**, 123108 (2019).

[23] H. Fan, J. Jiang, C. Zhang, X. Wang, and Y.-C. Lai, Phys. Rev. Res. **2**, 012080(R) (2020).

[24] A. V. Andreev, A. A. Badarin, V. A. Maximenko, and A. E. Hramov, Chaos **32**, 103126 (2022).

[25] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Phys. Rev. Lett. **120**, 024102 (2018).

[26] J. A. Platt, S. G. Penny, T. A. Smith, T.-C. Chen, and H. D. Abarbanel, Neural Netw. **153**, 530 (2022).
[27] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, Chaos **27**, 121102 (2017).
[28] M. Roy, S. Mandal, C. Hens, A. Prasad, N. V. Kuznetsov, and M. Dev Shrimali, Chaos **32**, 101104 (2022).
[29] S. Saha, A. Mishra, S. Ghosh, S. K. Dana, and C. Hens, Phys. Rev. Res. **2**, 033338 (2020).
[30] J. A. Platt, A. Wong, R. Clark, S. G. Penny, and H. D. Abarbanel, Chaos **31**, 123118 (2021).
[31] A. E. Hramov, A. A. Koronovskii, and O. I. Moskalenko, Phys. Lett. A **354**, 423 (2006).
[32] A. E. Hramov, N. Kulagin, A. V. Andreev, and A. N. Pisarchik, Chaos Solit. Fractals **178**, 114354 (2024).
[33] E. M. Izhikevich, *Dynamical Systems in Neuroscience* (MIT Press, Cambridge, MA, 2007).
[34] A. N. Pisarchik and A. E. Hramov, Phys. Rep. **1000**, 1 (2023).
[35] A. N. Pisarchik, V. A. Maksimenko, A. V. Andreev, N. S. Frolov, V. V. Makarov, M. O. Zhuravlev, A. E. Runnova, and A. E. Hramov, Sci. Rep. **9**, 18325 (2019).
[36] P. E. Kloeden, E. Platen, P. E. Kloeden, and E. Platen, *Stochastic Differential Equations* (Springer, New York, 1992).
[37] A. S. Pikovsky and J. Kurths, Phys. Rev. Lett. **78**, 775 (1997).
[38] L.-W. Kong, Y. Weng, B. Glaz, M. Haile, and Y.-C. Lai, Chaos **33**, 033111 (2023).
[39] A. van Kan, A. Alexakis, and M.-E. Brachet, Phys. Rev. E **103**, 052115 (2021).
[40] S. Aumaître, K. Mallick, and F. Pétrélis, J. Stat. Phys. **123**, 909 (2006).
[41] V. Pyragas and K. Pyragas, Phys. Lett. A **384**, 126591 (2020).
[42] P. Landa, A. Zaikin, M. Rosenblum, and J. Kurths, Chaos Solit. Fractals **9**, 157 (1998).
[43] A. Campos-Mejía, A. Pisarchik, and D. Arroyo, Chaos Solitons Fractals **54**, 96 (2013).
[44] O. I. Moskalenko, A. A. Koronovskii, A. O. Selskii, and E. V. Evstifeev, Chaos **31**, 083106 (2021).
[45] A. A. Koronovskii, O. I. Moskalenko, and A. O. Selskii, Phys. Rev. E **109**, 064217 (2024).