RESEARCH ARTICLE | MARCH 19 2025

Strong and weak prediction of stochastic dynamics using reservoir computing

Special Collection: Nonlinear Dynamics of Reservoir Computing: Theory, Realization and Application

Alexander E. Hramov ; Nikita Kulagin ; Alexander N. Pisarchik ; Andrey V. Andreev

Chaos 35, 033140 (2025)
https://doi.org/10.1063/5.0252908

Strong and weak prediction of stochastic dynamics using reservoir computing

Cite as: Chaos 35, 033140 (2025); doi: 10.1063/5.0252908
Submitted: 12 December 2024 · Accepted: 5 March 2025 ·
Published Online: 19 March 2025

Alexander E. Hramov,[1,a)] Nikita Kulagin,[1] Alexander N. Pisarchik,[1,2] and Andrey V. Andreev[1]

AFFILIATIONS

[1]Baltic Center for Neurotechnology and Artificial Intelligence, Immanuel Kant Baltic Federal University, Kaliningrad, Russia
[2]Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain

Note: This paper is part of the Special Topic on Nonlinear Dynamics of Reservoir Computing: Theory, Application and Realization.
[a)]**Author to whom correspondence should be addressed:** hramovae@gmail.com

ABSTRACT

We propose an approach to replicate a stochastic system and forecast its dynamics using a reservoir computing (RC). We show that such machine learning models enable the prediction of the behavior of stochastic systems in a wide range of control parameters. However, the quality of forecasting depends significantly on the training approach used for the RC. Specifically, we distinguish two types of prediction—weak and strong predictions. We get what is called a strong prediction when the testing parameters are close to the training parameters, and almost a true replica of the system trajectory is obtained, which is determined by noise and initial conditions. On the contrary, we call the prediction weak if we can only predict probabilistic characteristics of a stochastic process, which happens if there exists a mismatch between training and testing parameters. The efficiency of our approach is demonstrated with the models of single and coupled stochastic FitzHugh–Nagumo oscillators and the model of an erbium-doped fiber laser with noisy diode pumping. With the help of a RC, we predict the system dynamics for a wide range of noise parameters. In addition, we find a particular regime when the model exhibits switches between strong and weak prediction types, resembling probabilistic properties of on–off intermittency.

Machine learning has become an indispensable tool in our contemporary data-driven world, and its application often involves data classification, uncovering hidden patterns, and, more recently, predicting the behavior of complex nonlinear systems using advanced machine learning methods, such as recurrent neural networks and their variants, which include reservoir computers. In the context of reservoir computing (RC), researchers have proposed numerous approaches to predict the behavior of dynamical systems exhibiting chaotic regimes, spatial complexity, and other intricate characteristics. A pressing challenge arises when dealing with stochastic systems in which dynamics are significantly influenced by noise. A canonical example of this challenge is the phenomenon of stochastic and coherent resonances, observed in nonlinear systems subjected to noise. These effects, prevalent in systems of diverse nature, such as neural networks, brain, lasers, and climate dynamics, have particular relevance to living organisms, potentially playing a crucial role in the recognition of weak signals by sensory systems, the control of neuronal ensembles, and other biological processes. The ability to predict the behavior of such stochastic systems, particularly in response to variations in noise parameters, holds significant promise in biomedicine, intelligent control system design, brain–computer interface development, and the implementation of adaptive biological feedback mechanisms. This paper presents an approach for predicting the behavior of stochastic systems using a specifically modified reservoir computer capable of effectively predicting the effects of coherent resonance. We distinguish between two types of predictions using the reservoir computer: strong prediction, which allows us to predict, with a given accuracy, the trajectory of a stochastic process under a specified noise level, and weak prediction, which estimates only statistical characteristics of the stochastic process. The transition from strong to weak prediction follows a scenario inherent to on–off intermittency. The proposed concept exhibits a resemblance to the concept of generalized, or noise-induced, synchronization, where the establishment of generalized synchronization, as diagnosed by the auxiliary system method, mirrors the implementation of the strong prediction mode in the reservoir computer influenced by a set of noise signals. Furthermore, using a model of two coupled stochastic neurons as an example, we demonstrate that insufficient or

redundant information on noise sources significantly diminishes the accuracy of predicting the stochastic system behavior.

I. INTRODUCTION

We are surrounded by complex nonlinear systems of different nature and at different scales of observation—from genes and proteins to socio-economic and technogenic networks. As a rule, such systems in the process of their development in time demonstrate the formation of complex temporal or spatial patterns, called dissipative structures,[1] whose emergence is determined by internal nonlinear interaction, and their existence is stabilized by the exchange of matter, energy, or information with the surrounding world.[2–5] The description of such systems can rarely be achieved only analytically, so methods of numerical analysis and prediction of the behavior of complex non-equilibrium systems are of great interest. Moreover, researchers are increasingly considering systems whose mathematical models are either unknown or so complex that direct mathematical and numerical analysis is not possible. This is a reason why data-driven approaches that build models for poorly formalized processes and phenomena are gaining attention.[6,7] Large amounts of data on many systems' dynamics can be collected so that our predictions on the system behavior under study as well as certain decisions are based on the analysis and interpretation of collected data. Models based on machine learning (ML) methods[7–9] cope most effectively with such a task as predicting the behavior of various systems based on big data. The ML-based models effectively generalize features by relying on carefully trained and annotated datasets containing empirical, often experimental, data about the system under study.

Among various data-driven and model-independent ML methods, reservoir computing (RC)[10] has recently gained widespread popularity due to its effectiveness in predicting and classifying dynamical systems.[2,11–14] A RC is a type of recurrent neural network architecture designed for various ML and multivariate time series prediction tasks. The effectiveness and appeal of RCs lie in the simplicity of their architecture, strong predictive capabilities, and cost-effective learning by requiring only the output layer of the reservoir to be trained. Recently, the RC has played a key role in predicting a variety of dynamical systems' characteristics, including predicting chaotic time series,[15,16] quantifying chaos,[17,18] and predicting cluster and burst synchronization,[19,20] observation of turbulent spatiotemporal dynamics,[21,22] characterization of macroscopic properties of complex networks,[23] modelling of basins of attraction,[24] etc.

It is important to note that the behavior of complex non-equilibrium systems is often determined not only by dynamical processes, but also by stochastic influences or processes in the systems.[25–27] Stochastic processes can often lead to new nonlinear effects, such as stochastic[28,29] and coherent[30] resonance, noise-induced bifurcations and intermittency,[31–33] noise-induced second-order phase transitions,[34,35] noise-enhanced stability,[36] etc. In many fields of natural science, we are faced with recording or collecting some empirical data that are more or less disordered signals, images, or data. The interpretation of such observations often brings uncertainty, which in turn requires the involvement of probability in their description. Consequently, the construction of their mathematical models leads to stochastic differential equations or stochastic maps with random variables.[37] In the case of impossibility to build such models, data-driven models come to the fore, allowing to use the accumulated empirical data to build ML-based models, including those using RC.

However, it should be noted that there is a certain gap in the RC application for predicting the dynamics of stochastic systems. Recent studies in this area include the work of Grigoryeva and colleagues,[38] in which a time-delayed RC demonstrated robust performance in predicting the conditional covariance associated with multivariate discrete nonlinear stochastic processes of the VEC-GARCH type. This RC-based model also performed well in predicting the actual daily realized market volatility based on intraday quotes, using daily log-return series of modest size as training inputs. In the work of Fang *et al.*,[39] a data-driven model combines RC and normalizing flow to predict the long-term evolution of stochastic dynamical systems and reproduce their behavior. The authors validate the effectiveness of this framework through various simulations covering systems, such as the stochastic Van der Pol and Lorenz oscillators and the simulated El Niño–Southern Oscillation model. Recently, Liao *et al.*[40] has developed a low-power physical RC-based model based on a bistable stochastic resonant system with excessive damping, offering an innovative path for efficient computing in stochastic environments. In addition, Hramov *et al.*[41] have proposed an RC design for predicting dynamics of the stochastic FitzHugh–Nagumo (SHFN) neuron, which turns out to be very effective over different ranges of the noise control parameters, exactly replicating the main characteristics of the original stochastic neuron, including the coherence resonance effect.

However, it should be concluded that there are currently no effective frameworks for reservoir computing capable of accurately predicting stochastic systems, and there are only a few successful examples of private problem analyses. This highlights the critical importance of investigating and predicting the dynamics of systems affected by noise and makes the task of building reservoir computing frameworks for predicting stochastic systems relevant and significant. At the same time, it is important to understand how we should address such a challenge, in particular, whether we should set the problem of accurate prediction of the dynamics of a stochastic system using reservoir computing or whether it is sufficient to predict some important characteristics of the stochastic process for a particular problem. The first approach, which we call *strong prediction*, of accurate prediction of the trajectory of a stochastic system in phase space is natural for a number of applied problems, in particular, for problems of optimal control theory or chaos control. However, for many tasks, the forecasting of even certain characteristics of a stochastic process or *weak prediction* is of primary interest. For example, to predict stochastic resonance, we do not need an exact copy of the trajectory of the stochastic system, but we need to accurately forecast a particular coherence measure of the stochastic system dynamics,[30] which can be assumed to be a simpler task than an exact prediction of the trajectory.

Usually, in the context of predicting the dynamics of nonlinear systems, it is expected to obtain an accurate prediction of the trajectory of the system. For example, in the case of chaotic systems,

this is possible only on some time interval defined by the largest Lyapunov exponent of the chaotic system.[15,23] In our terminology, this corresponds to a strong prediction, that is, to an accurate prediction of the trajectory using RC, only at some time interval after the start of the prediction. However, in the study of chaotic systems, the problem of predicting the characteristics of the dynamics of chaotic systems, for example, the Lyapunov exponents, is posed,[17] and in this case, the problem is reduced to the reproduction of the chaotic attractor of the system with the help of the RC without setting the problem of exact reproduction of the system trajectory.[42] Such a problem can be regarded as an analog of weak prediction in our case, i.e., prediction of statistical characteristics of a stochastic process. Thus, standard approaches to predicting the temporal dynamics of systems using RC usually correspond to the case of strong prediction in our terminology. The models that provide weak predictions may be simpler than the corresponding strong predictive models of stochastic processes since the only goal of the weak prediction models is to forecast the probability distribution. Such models may be useful, for example, for building financial models when the goal is to predict the probability distribution law of an asset price or interest rate,[43] or in biology to predict collective dynamics of population processes[44] or neural ensembles.[27] These applications are mainly generated by the interest in finding one of the statistical moments of a stochastically defined variable, and ML-models for weak prediction may be simpler and sufficient for the chosen purpose.

Thus, in this paper, we introduce the notion of strong and weak predictions of the dynamics of stochastic systems using reservoir computing and consider the relation between weak and strong predictions using as examples the stochastic FitzHugh–Nagumo neuron, two coupled stochastic FitzHugh–Nagumo neurons, and the model of an erbium-doped fiber laser (EDFL) with noisy diode pumping.

II. GENERAL FORMALISM

As we already mentioned in Sec. I, there are many dynamical models described by differential equations where random fluctuations should properly be introduced. "How to do this" is one of the most frequent problems in different areas of science (radio engineering, thermodynamics, kinetics of chemical reactions, neural ensembles, astrophysics, etc). One way is the use of the Langevin equation,

$$\dot{x}_i = f_i(\mathbf{x}) + \sum_{\alpha=1}^{m} g_i^{\alpha}(\mathbf{x})\xi_{\alpha}(t), \quad i = 1, \ldots, k, \tag{1}$$

where $\mathbf{x} = \{x_i | 1 \le i \le k\}$ is the set of unknown functions, f_i and g_i are some functions describing a particular problem, and ξ_{α} are random functions of time. If $g_i(\mathbf{x}) \propto \mathbf{x}$, then we are dealing with a system with multiplicative noise. If g_i are constants, then the system is said to be subject to additive noise. In this paper, we will consider the simpler case of additive noise.

From the point of view of mathematics, stochastic systems are usually described in terms of some random process, i.e., noise models. The most common noise is white noise described by means of the Wiener process $W(t)$, $t > 0$. This is a Gaussian process with continuous trajectories, zero mean $\mathbb{E}W(t) = 0$, $t > 0$, and covariance function $K(s,t) = \mathbb{E}W(s)W(t) = \min(s,t)$. In other words, this is a continuous process with independent homogeneous increments, with $W(0) = 0$. The distribution of increments $W(s) - W(t)$ is normal with zero mean and variance $(s - t)$. To describe a multivariate stochastic process, the corresponding system of stochastic differential equations can be written in a general form as

$$d\mathbf{X}(t) = \mathbf{a}(t, \mathbf{X}(t))dt + \sigma(t, \mathbf{X}(t))d\mathbf{W}(t), \quad t \ge 0, \tag{2}$$

with initial conditions $\mathbf{X}(0)$ as random variables independent with process $\mathbf{W}(t)$.

In Eq. (2), variable $\mathbf{X}(t)$ and drift $\mathbf{a}(t, \mathbf{X}(t))$ are k-component column vectors, $\mathbf{W}(t)$ is a m-component column vector, and $\sigma(t, \mathbf{X}(t))$ is a $k \times m$ matrix. There are m sources of noise, modeled by the m-component Wiener process $\mathbf{W}(t)$ (i.e., the components of $\mathbf{W}$ are independent one-dimensional Wiener processes). It should be noted that equality $k = m$ is not required, and the case $m < k$ is common in applications. Equation (2) can be considered a reduced form of the corresponding equation in an integral form for the ith component of the column vector $\mathbf{X}$ as

$$X_i(t) = X_i(0) + \int_0^t a_i(s, \mathbf{X}(s))ds + \int_0^t \sum_{\alpha=1}^{m} \sigma_{i,\alpha}(s, \mathbf{X}(s))dW_{\alpha}(s), \quad i = 1, \ldots, k, \tag{3}$$

where the second integral can be understood in the Ito sense or in the Stratonovich sense, which is determined by the convenience of analysis and qualitative properties of the analyzed system.

When posing the problem of predicting behavior of the stochastic system [Eq. (1)] in the case of additive noise using ML, we must keep in mind a number of factors. ML involves generating a data set that describes the system being predicted to some extent and allows us to train a predictive model based on it. Therefore, the prediction quality will crucially depend on the completeness of the collection of information about the behavior of the predicted system. For example, the success of reservoir computing for forecasting dynamical systems suggests that this is a solvable problem in the case of a system not affected by noise (or such an effect is not critical for making short-term or long-term forecasts so that noise effects can be neglected). Therefore, when training RC, we do not use any data on noise effects, restricting ourselves only to data on the dynamical variables of the predicted system.

However, for stochastic systems, we fundamentally need to take into account the stochastic impact factor in the predictive model. In Figs. 1(a) and 1(b), we present a general scheme of RC training in the classical case without noise influences and in a stochastic system, respectively. One can clearly see two important effects of noise in the latter case. First, when generating the training dataset, we need to collect noise affect data in addition to the dynamical variables. Second, even in the case of additive noise, we cannot always ensure precise knowledge of all stochastic factors influencing the system.

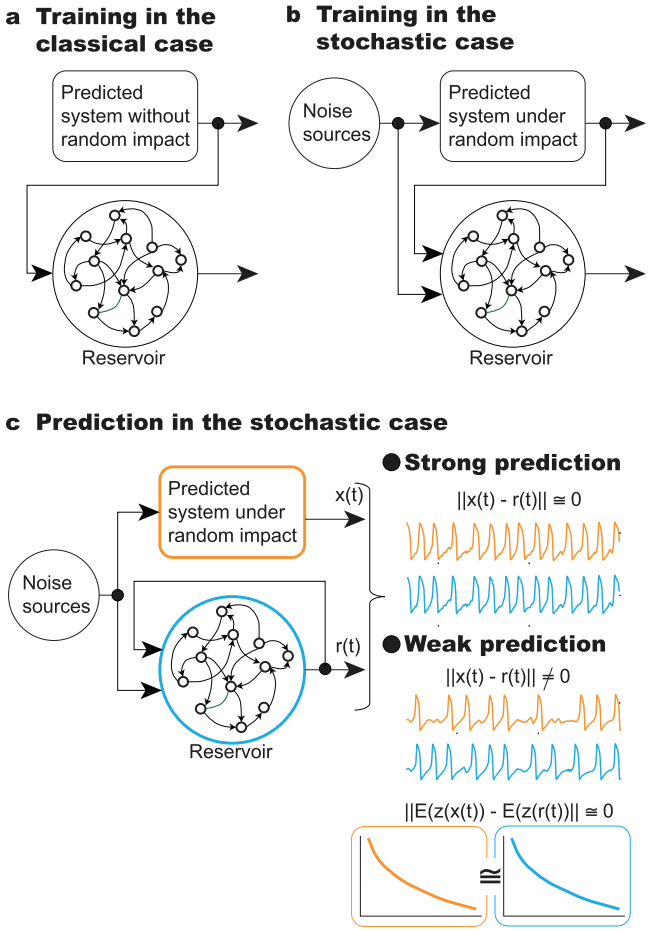

FIG. 1. General schemes for RC training and RC prediction of the stochastic systems dynamics. (a) RC training in the classical case in the absence of noise effects. For RC training, only a dataset of trajectories of the system under study at some time interval is needed. (b) RC training in the case of a stochastic system. In this case, the RC training requires not only the dataset of trajectories of the system under study at time interval δT, but also the knowledge of noise influences during time δT. (c) Prediction using a RC trained on data on the behavior of dynamical variables and noise. To evaluate the accuracy of the prediction, we have to expose the predicted system and the RC to the same noise and compare the output characteristics of the stochastic system and the RC. In the case of strong prediction, we will get an accurate prediction of the system trajectory, while in the case of weak prediction, only a match of the statistical characteristics.

As a result, we often rely on hypotheses regarding the nature of these stochastic influences, which means that accurate prediction of the system's trajectory cannot be guaranteed. However, assuming or knowing the statistical properties of the noise impact, we can solve a less ambitious problem; namely, we can try to predict the statistical characteristics of the stochastic system's dynamics, rather than its exact trajectory.

The above observations lead us to conclude that the concept of predicting the dynamics of stochastic systems using ML-models requires classification. Specifically, it is important to distinguish between strong and weak predictive models, as illustrated in Fig. 1(c). Suppose that our stochastic system is described by a model based on the stochastic differential equation (2). To predict the behavior of the stochastic system, we need to train a ML-model, for which we usually solve the stochastic differential equation (2) numerically.

When we solve a stochastic system (2) in Ito formalism, we employ the Euler–Maruyama (EM) numerical method[45] for discrete time points $t = 0, \Delta t, 2\Delta t, \ldots, n\Delta t, \ldots, L\Delta t$, where $n \in \mathbb{N}$ and $0 \le n \le L$, Eq. (2) is replaced with

$$X_{i,t+\Delta t} = X_{i,t} + a_i(t, \mathbf{X}_t)\Delta t + \sum_{\alpha=1}^{m} \sigma_{i,\alpha}(t, \mathbf{X}_t)\Delta W_{\alpha,t}, \tag{4}$$

$$\Delta \mathbf{W}_t = \mathbf{W}(t + \Delta t) - \mathbf{W}(t), \quad \mathbf{X}_t = \mathbf{X}(t), \quad t = n\Delta t,\ i = 1, \ldots, k. \tag{5}$$

In the case of Stratonovich formalism, we should use the Euler–Heun (EH) numerical method,[46] which requires calculation of predictor and corrector steps,

$$\tilde{X}_{i,t+\Delta t} = X_{i,t} + a_i(t, \mathbf{X}_t)\Delta t + \sum_{\alpha=1}^{m} \sigma_{i,\alpha}(t, \mathbf{X}_t)\Delta W_{\alpha,t}, \tag{6}$$

$$X_{i,t+\Delta t} = X_{i,t} + \frac{1}{2}\left[a_i(t, \mathbf{X}_t) + a_i(t + \Delta t, \tilde{\mathbf{X}}_{t+\Delta t})\right]\Delta t + \frac{1}{2}\sum_{\alpha=1}^{m}\left[\sigma_{i,\alpha}(t, \mathbf{X}_t) + \sigma_{i,\alpha}(t + \Delta t, \tilde{\mathbf{X}}_{t+\Delta t})\right]\Delta W_{\alpha,t}, \tag{7}$$

$$\Delta \mathbf{W}_t = \mathbf{W}(t + \Delta t) - \mathbf{W}(t), \quad \mathbf{X}_t = \mathbf{X}(t), \quad t = n\Delta t,\ i = 1, \ldots, k. \tag{8}$$

When numerically solving the SDEs, we calculate ΔW_t as $\xi_t\sqrt{\Delta t}$, where ξ is the zero-mean white Gaussian noise with standard deviation.

Through training, the model learns to predict a stochastic process using the trajectories of the process $\mathbf{X}_t$ and the external noise sources $\mathbf{W}_t$, as shown in Fig. 1(b). In a prediction mode, our predictive model generates a signal $\mathbf{r}_t$, which serves as the prediction and must be compared in some manner with the true target process $\mathbf{X}_t$.

In the case of strong prediction, it is assumed that the ML-based predictive model enables the determination of the exact trajectory of the stochastic system's dynamics (2). This trajectory is fully specified at time t by the given trajectory of the Wiener process $\mathbf{W}$ over the temporal interval $[0, L\Delta t]$ and the initial condition $\mathbf{X}_0$,

$$\max_{t=0,\ldots,L\Delta t} \mathbb{E}\left(|\mathbf{X}_t - \mathbf{r}_t|\right) \le \epsilon, \tag{9}$$

where $\mathbf{r}_t$ is the process predicted by the ML-based model, $\mathbf{X}_t$ is the true target process determined by the EM numerical model (4)–(5)

or the EH numerical model (6)–(8) of the stochastic system (2), and ϵ denotes the prediction accuracy—a small value that determines the quality of a strong predictive model.

In the second case, a weak predictive model implies the qualitative construction of a probabilistic model (probability space) where some Wiener process W and process X satisfying equation (2) are given. A predictive model gives a weak prediction if the condition

$$\max_{t=0,\ldots,L\Delta t} |\mathbb{E}(z(\mathbf{X}_t)) - \mathbb{E}(z(\mathbf{r}_t))| \leq \epsilon \tag{10}$$

is fulfilled for all polynomials $z(\cdot)$. In this case, all moments of distribution of stochastic processes $\mathbf{X}_t$ and $\mathbf{r}_t$ turn out to be the same with a given accuracy ε.

In other words, in the case of strong prediction, we reconstruct the trajectories of a pair of random processes $(\mathbf{W}, \mathbf{X})$ with a certain accuracy on some interval $[0, t]$, while in the case of weak prediction, we reconstruct only the distributions of this pair. The construction of the weak prediction model in a strong sense automatically entails the construction of the prediction in a weak sense.

Thus, depending on the nature of the predictive problem to be solved, as well as on the data set available for training the predictive ML-based model, we can forecast strongly or weakly the behavior of a stochastic system. In the former case, an approximate trajectory of the Wiener process is constructed from the specific trajectory of the process available in the dataset; i.e., the closeness of the true and predicted trajectories can be understood in terms of mean square, probability (uniform or point-wise), convergence of approximations, etc. However, as we have already noted, for many applications, a precise knowledge of the future trajectory is not necessary, and a weak prediction of probabilistic properties is enough. From the mathematical point of view, if we return to the stochastic equation (2), this means the approximation of the joint distribution of the Wiener process and the prediction of the solution of the stochastic system. The closeness can be understood as the closeness of finite-dimensional distributions or in the sense of weak convergence of probability measures on the corresponding functional space of trajectories.[47] In other words, the problem of weak stochastic prediction can be solved by realizing a large number of N independent trajectories of a process close in distribution to the real observable process. In the case of an ergodic system, this means that we can predict the system behavior in time and estimate the statistical characteristics of the forecast from a single time series generated by the model.

To investigate the considered approach of separating strong and weak prediction of stochastic systems, we take into account three benchmark models that exhibit complex stochastic dynamics under varying control parameters and noise power [see details in Sec. III A]: (i) the stochastic FitzHugh–Nagumo neuron (SFHN) demonstrating the effect of stochastic resonance when varying noise intensity;[48] (ii) the coupled SFHN neurons as a model demonstrating the limits of predictability of interaction effects of two stochastic systems;[49,50] and (iii) the noise-pumped erbium-doped fiber laser (EDFL) model[51,52] as an example of an excitable physical system under noise influence.

III. METHODS

A. Stochastic systems under study

As the first mathematical model under study, we consider the SFHN model given by the following system of equations with additive noise term $D\xi(t)$:

$$\begin{aligned} \dot{x} &= x - x^3/3 - y + 0.3, \\ \dot{y} &= 0.08(x - 0.8y + 0.7) + D\xi(t), \end{aligned} \tag{11}$$

where x and y represent the excitatory and recovery variables, respectively, $\xi(t)$ is the zero-mean white Gaussian noise with an autocorrelation function of $\langle \xi(t)\xi(t+t_0)\rangle = \delta(t_0)$, and D represents the noise amplitude.

In the deterministic case ($D = 0$), the SFHN model given by Eq. (11) represents a steady-state solution. However, as the noise amplitude D is increased, the system (11) exhibits a spiking behavior with interspike intervals (ISIs) dependent on the noise intensity. In addition, such a system demonstrates the effect of coherence resonance,[30] which consists in maximizing regularity of the generated spike sequences at a certain optimal noise amplitude, in our case at $D \approx 0.2$.

For the stochastic FHN neuron (11), the RC for predicting signals $(x(t), y(t))$ was constructed in Ref. 41 where the possibility to predict the effect of coherence resonance was demonstrated by varying the noise amplitude D.

To compare the effect of different numerical approaches to solving the stochastic equations (11) on observed RC prediction quality, we employed both the EM method (4) and (5) and the EH method (6)–(8) with a time step-size set at $\Delta t = 0.1$. Looking ahead, we note that the prediction results in our case are practically independent of the numerical solution scheme (see Sec. IV A). Therefore, in the following, we use only the EM numerical method.

Another stochastic system which we will consider in this paper is two coupled SFHN neurons. Similar to Refs. 49 and 50, we consider electrically coupled SFHN neurons and analyze the limit of predictability of interaction effects of two stochastic systems. The corresponding model is written in the following form:

$$\begin{aligned} \dot{x}_1 &= x_1 - x_1^3/3 - y_1 + 0.3 + \rho(x_2 - x_1), \\ \dot{y}_1 &= 0.08(x_1 - 0.8y_1 + 0.7) + D_1\xi_1(t), \\ \dot{x}_2 &= x_2 - x_2^3/3 - y_2 + 0.3 + \rho(x_1 - x_2), \\ \dot{y}_2 &= 0.08(x_2 - 0.8y_2 + 0.7) + D_2\xi_2(t), \end{aligned} \tag{12}$$

where ρ is the coupling strength and $\xi_1(t)$ and $\xi_2(t)$ are two different zero-mean white Gaussian noises with the same distribution and autocorrelation function of $\langle \xi(t)\xi(t+t_0)\rangle = \delta(t_0)$.

The system of two coupled SFHN neurons, represented by Eq. (12), operates in an excitable regime similar to that of the model described in Eq. (11). The key distinction lies in the fact that when $D \neq 0$, each SFHN neuron generates spikes based on its own intrinsic noise and the coupling between them. This model also demonstrates the phenomenon of coherence resonance; however, the optimal noise amplitude D is contingent upon the coupling strength ρ—an increase in ρ corresponds to a higher optimal value

of D. For numerical solutions of Eq. (12), we employed the EM method with a time step-size set at $\Delta t = 0.1$.

Finally, as an example of a physical system that can also operate in the excitable regime and demonstrate stochastic dynamics under the influence of a noise signal, consider the model of a 1560 nm EDFL with a Fabry–Pérot cavity that has been subjected to noisy pump modulation of a diode pumping laser proposed in Ref. 51. The EDFL model equations have the following form in dimensionless variables:

$$\dot{x} = axy - bx + c(y + \upsilon),$$
$$\dot{y} = -dxy - (y + \upsilon) + P_p \left\{1 - \exp\left[-18\left(1 - \frac{y + \upsilon}{\rho}\right)\right]\right\}, \quad (13)$$

where x and y are the intracavity laser power and averaged over the active fiber length population of the upper level system, respectively, $P_p = P_{p0}(1 + D\xi(t))$ is the pump parameter, $P_{p0} = 506$ is the pump power without modulation ($D = 0$), and $\xi(t)$ is the zero-mean white Gaussian noises. The following parameters were used:[52] $a = 6.6207 \times 10^7$, $b = 7.4151 \times 10^6$, $c = 0.0163$, $d = 4.0763 \times 10^3$, $\upsilon = 0.3075$, and $\rho = 0.6150$. These corresponded to experimentally evaluated parameters of a real EDFL with a 70-cm active erbium-doped fiber. With the chosen parameters, the deterministic EDFL (13) operates in a steady-state regime similar to the SFHN neuron (11). As the noise amplitude is increased, the noise-pumped EDFL generates noise-induced oscillations of $x(t)$ and $y(t)$ variables. For numerical solutions of Eq. (13), we employed the EM method with a time step-size set at $\Delta t = 10^{-7}$.

B. RC-based model for prediction of stochastic system dynamics

1. Architecture of the RC

In Ref. 41, a RC design was proposed for predicting the SFHN neuron dynamics described by the stochastic differential equation with a single noise source. Let us generalize this ML-model to predict the behavior of stochastic systems with a large number of noise sources.

The RC architecture includes three main components: (i) the input layer, which receives input data in the form of time series which visualize temporal dynamics of the predicted system and feeds the input data to the reservoir layer, (ii) the reservoir layer, which consists of a large number of artificial neurons connected by randomly generated links, and (iii) the output layer connected to the neurons of the reservoir layer and whose connection weights are tuned for prediction based on the information stored and processed in the reservoir layer.

Figure 2 illustrates the proposed RC configuration. In the case of modeling a stochastic system, it is necessary to provide reservoir inputs that correspond to noise sources that determine the system dynamics. In the output layer, we should not provide corresponding outputs to predict noise processes, which is impossible due to their nature. Therefore, when predicting the dynamics of stochastic systems, we always need to know times series of noise that affects the system behavior in order to build a strong prediction model. In this case, in the input layer, we have K inputs corresponding to the dynamical variables $\mathbf{X}_t = \left(X_{1,t}, X_{2,t}, \ldots, X_{K,t}\right)^T$ and M inputs describing either additive or multiplicative random fluctuations $\left(\xi_{1,t}, \xi_{2,t}, \ldots, \xi_{M,t}\right)^T$. Following the paper,[41] a distinctive aspect of such proposed RC configuration is the complete segregation of reservoir inputs among different artificial neurons within the hidden inner layer. In other words, each neuron in the reservoir layer is associated with only one input in the input layer. For simplicity, we can assume that each input is associated with the same number of neurons in the reservoir layer, namely, with $N_h/(K + M)$, where N_h is the number of artificial neurons in the reservoir layer.

In general, we may have no information about some number M_0 of noise sources. In this case, our predictive model will include only $(M - M_0)$ inputs with noise, which leads to the question about the possibility of weak predicting the behavior of the system. At the same time, we may get a situation where some noise impacts are mistakenly considered important in analyzing the system. Then, M_e of erroneous noise sources may appear additionally as inputs, which we measure in the experiment, but which do not play an essential role in the dynamics of the stochastic system. We will address this question further when discussing the results of the prediction analysis.

As a consequence, the input $(K + M - M_0 + M_e)$-component column vector denoted as $\mathbf{g}_t$ includes $(M - M_0 + M_e)$ noise components and K dynamical variables, namely,

$$\mathbf{g}_t = \left(\xi_{1,t}, \xi_{2,t}, \ldots, \xi_{M-M_0+M_e,t}, X_{1,t}, X_{2,t}, \ldots, X_{K,t}\right)^T, \quad (14)$$

and the output K-component column vector denoted as $\mathbf{r}_{t+1}$ has only K components corresponding to predicted dynamical variables, namely,

$$\mathbf{r}_{t+1} = \left(r_{1,t+1}, r_{2,t+1}, \ldots, r_{K,t+1}\right)^T. \quad (15)$$

Each of N_h neurons of the reservoir layer receives $\mathbf{g}_t$ signals from the input layer as well as signals from other neurons in the reservoir, which are transformed according to the following equation:[41]

$$\mathbf{h}_{t+1} = \tanh(\mathbf{H}\mathbf{h}_t + \mathbf{G}\mathbf{g}_t), \quad (16)$$

where $\mathbf{h}_t$ represents the internal hidden state at the time moment t, allowing for the encoding of temporal dependencies based on past state history and input data (14). $N_h \times N_h$ matrix $\mathbf{H}$ is the reservoir's adjacency matrix, which sets the weights of connections between artificial neurons of the reservoir layer. This matrix defines a random network characterized by the average node degree $\bar{k}$ and the spectral radius λ (absolute value of the largest eigenvalue of the adjacency matrix $\mathbf{H}$). $(K + M - M_0 + M_e) \times N_h$ matrix $\mathbf{G}$ is the input matrix defining the coupling of the input to the hidden state $\mathbf{h}$. In the case of the above condition of complete segregation of reservoir inputs among different artificial neurons within the reservoir layer, it is necessary to choose the number of artificial neurons in the reservoir layer such that the ratio $c = (K + M - M_0 + M_e)/N_h$ is integer, and each resevoir input affects c artificial neurons in the reservoir layer. Then, we can set the elements of the input matrix $\mathbf{G}$ as $G_{ij} = \eta_{i,j}$ if $1 + (j - 1)c \le i \le jc$ and $G_{ij} = 0$ in other cases, where $i = 1, \ldots, (K + M - M_0 + M_e)$, $j = 1, \ldots, N_h$, $\eta_{i,j}$ are uniformly sampled from interval $[-\sigma, \sigma]$.

The goal of RC in the considered case is to predict the behavior of the stochastic process under study $\mathbf{r}$ [see Eq. (15)] based on the observed signals of dynamical variables and noise sources $\mathbf{g}$ [see Eq. (14)] during time interval δT. Using the matrix $\mathbf{R}$ specifying the

weights of the output layer, we compute the predicted signal $\mathbf{r}$ at time $(t+1)$ by the following equation:

$$\mathbf{r}_{t+1} = \mathbf{R}\hat{\mathbf{h}}_t, \tag{17}$$

where $\hat{\mathbf{h}}$ is the augmented reservoir state represented as an N_h component column vector with components $h_{i,t} = h_{i,t}$ if i is odd, and $h_{i,t} = h_{i,t}^2$ if i is even, $i = 1, \ldots, N_h$[24].

2. Training mode of the RC

In the training mode, we must determine the output layer matrix $\mathbf{R}$. For this purpose, as shown in Fig. 2(a), the input vector $\mathbf{g}_t$, including both driving noise sources and dynamical variables $\mathbf{x}_t$, is fed to the input of the reservoir and the corresponding output signal $\mathbf{r}_{t+1}$ is obtained. To find the coefficients of the $\mathbf{R}$ matrix, we minimize L_2–error,

$$L_2 = \sum_{t=1}^{\delta T} \|\mathbf{r}_t - \mathbf{X}_t\|^2 + q\|\mathbf{R}\|^2, \tag{18}$$

between target state $\mathbf{X}$ and predicted state $\mathbf{r}$. The second term in Eq. (18) is included to prevent the overfitting problem, the regularization parameter is denoted by $q = 10^{-9}$, and δT is the duration of the dataset $\mathbf{g}$ that was used to train the RC.

3. Predicting mode of the RC

In the predicting mode, we use the RC in which input, inner, and outer layers are described by Eqs. (16) and (17), where the $\mathbf{G}$ and $\mathbf{H}$ matrices are the same as in the training mode, and the matrix $\mathbf{R}$ of the output layer is determined in the training phase by minimizing the L_2–error (18). In this mode, we feed the predicted values of the dynamical variables from the reservoir output to the reservoir input, as shown in Fig. 2(b); that is, the values of $\mathbf{r}_{t+1}$ are assigned to the values of $\mathbf{X}_{t+1}$, which are fed back to the reservoir input. However, according to our approach, which is illustrated in Fig. 1(c), we simultaneously feed those noise signals $(\xi_1, \xi_2, \ldots, \xi_{M-M_0+M_e})$ that affect the predicted stochastic system as well. Thus, unlike predicting the behavior of dynamical systems using a RC (see Fig. 1(a)), we cannot achieve full autonomy of the RC because we need data on noise sources.

4. Hyperparameter selection

Finally, let us focus on the choice of hyperparameters of the RC used for prediction. The reservoir layer size was chosen as $N_h = 500$ neurons. The reservoir hyperparameters were optimized by grid search in the following parameter ranges: spectral radius $\lambda \in [0.1, 1.9]$ and average node degree $\bar{k} \in [10, 20]$. The strength parameter of the input links was chosen to be $\sigma = 1.0$. After training, we obtain a set of 209 reservoirs, one for each $(\sigma, \bar{k})$ pair. The reservoirs were trained using time series of length $\delta T = 30\,000$ for SFHN neurons (11) and $\delta T = 40\,000$ for EDFL (13). As a result, the matrix of output weights $\mathbf{R}$ is formed by minimizing the L_2-error function (18).

The standard hyperparameter optimization strategy is to test all reservoirs on the same test signal and select the combination of hyperparameters that gives the lowest L_2–error. It is this reservoir that is used to further predict the behavior of the stochastic system. In this study, we trained the RC using this method to predict the dynamics of the SFHN model at noise amplitude $D_0 = 0.2$ and the EDFL model at noise amplitude $D_0 = 0.4$. Following the results of Ref. 41, we can assume that a reservoir trained at this noise value will be able to predict the dynamics of a stochastic system at other values of noise D.

However, in this paper, we added the following modification to the process of selecting the optimal set of reservoir layer hyperparameters. Our goal was to obtain the best prediction not only at a particular value of noise intensity D_0, but also at other values of noise. Therefore, when searching for the optimal set of hyperparameters, we tested each RC on a set of signals with three different values of the noise amplitudes, namely, $D = \{0.05, 0.2, 1.0\}$ for the SFHN model and $D = \{0.1, 0.4, 1.2\}$ for the EDFL model. Then, we chose the combination of hyperparameters that gives the smallest mean error over all three values of noise amplitudes D. This RC was used to predict the behavior of the stochastic system because it had the best generalizability of predicting the system dynamics for different noise amplitudes.

C. Evaluation of accuracy measures of strong and weak predictions

To assess the accuracy of the strong prediction, we need to compare the true trajectory of the system in phase space and the predicted trajectory made by the RC. First, we normalize the trajectories of all dynamical variables of both the true $\mathbf{x}$ and those $\mathbf{r}$ predicted by the RC as follows: $\bar{x}_i = (x_i - x_{i,0})/x_{i,\max}$ and $\bar{r}_i = (r_i - r_{i,0})/r_{i,\max}$, where $x_{i,0}$, $r_{i,0}$ are the mean values and $x_{i,\max}$, $r_{i,\max}$ are the maximum absolute values of the true and predicted variables over time interval ΔT within which the RC prediction was tested. Then, we use the root mean square error (RMSE), which is calculated along the normalized phase-space trajectory of the original stochastic system,

$$\mathrm{RMSE} = \sqrt{\frac{1}{\Delta T}\sum_{t=1}^{\Delta T} \|\bar{\mathbf{r}}_t - \bar{\mathbf{x}}_t\|^2}. \tag{19}$$

The RMSE takes a value from 0 (perfect prediction) to 1 (no prediction at all). As a measure of the accuracy of strong prediction, we use the value

$$\delta = 1 - \mathrm{RMSE}. \tag{20}$$

As we will see below, it is possible that strong prediction may not be achieved during the whole time δT, but only in some time intervals ΔT. To investigate time dependence of strong prediction, we calculate $\delta(t)$ in time window $[t, t + \Delta T]$ through the whole signal and transform it into a binarized form as follows:

$$\gamma(t) = \begin{cases} 1 & \text{if } \delta(t) > \delta_0, \\ 0 & \text{if } \delta(t) \le \delta_0. \end{cases} \tag{21}$$

Here, δ_0 is the threshold value characterizing the required accuracy of the system trajectory matching. In this study, we have chosen the threshold $\delta_0 = 0.95$. So, $\gamma(t) = 1$ corresponds to strong prediction in time interval $[t, t + \Delta T]$, while $\gamma(t) = 0$ means the lack of strong prediction on this interval. We will refer to $\gamma(t)$ as the quality dependence of the strong prediction on time during the RC predicting mode.

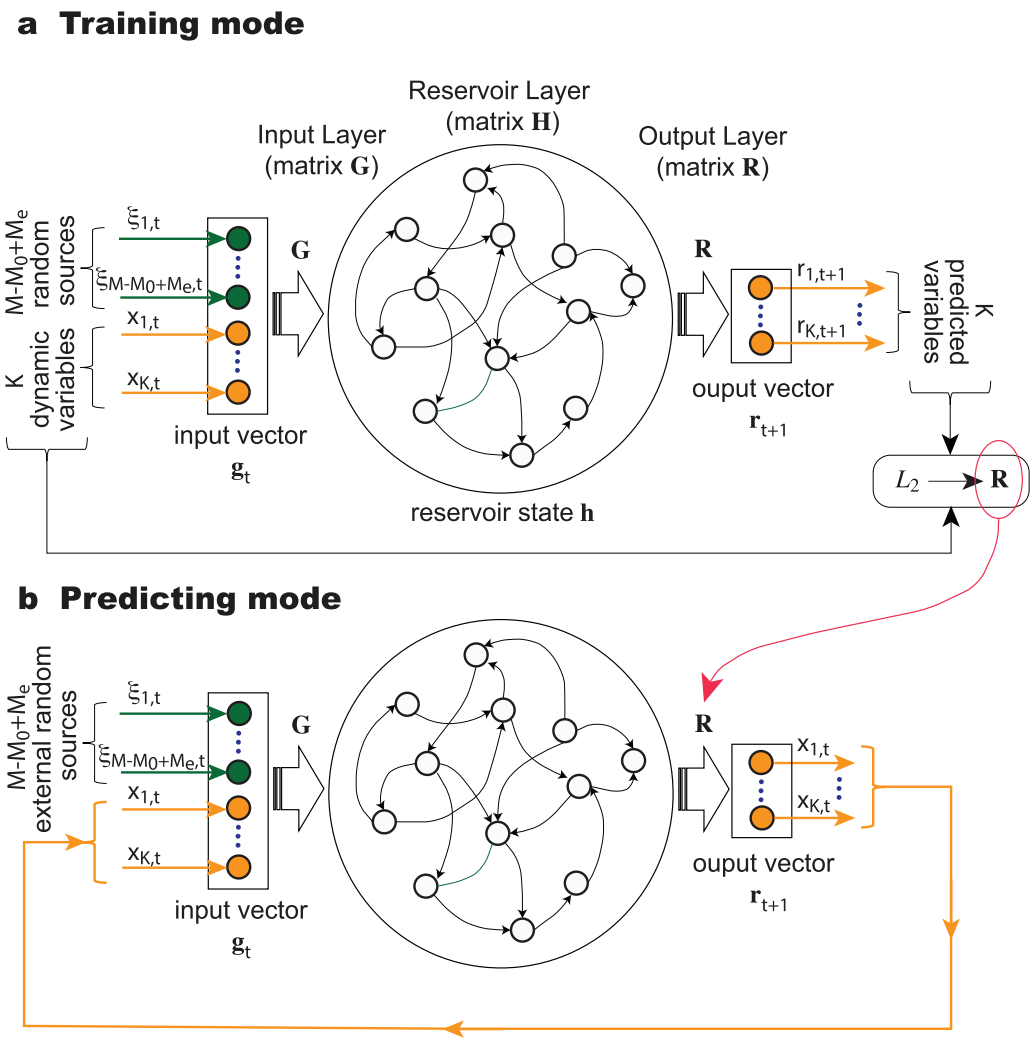

FIG. 2. Schematic presentation of the reservoir computing architecture for the stochastic systems dynamics prediction. (a) In a training mode, the signals (14) from noise sources and the signals generated by the stochastic system under study are fed to the input of the reservoir. The matrix of the output weights **R** is determined based on the L_2-error (18) minimization condition. (b) In a prediction mode, we use the output layer weight matrix found in the training mode, and according to the scheme in Fig. 1(c), we introduce feedback between the output and input of the RC by simultaneously feeding the signals of noise sources as an external influence on the RC.

To estimate the statistical characteristics of the spike train produced by the stochastic neuron and the RC-based model, we calculate the probability density function (PDF) of the duration of ISI $\Delta I_i = I_{i+1} - I_i$, where I_i is the time moment of the ith spike generation.[53] To analyze the statistical properties of the EDFL model, we consider the PDF of the laser power $x(t)$ [see Eq. (13)].

To analyze and compare the PDFs of the true stochastic dynamics of the SDEs under study and the dynamics predicted by the RCs, we use the coefficient of variation (CV) defined as the standard deviation σ_v of the analyzed variable v_i normalized to its average value $\langle v \rangle$,[30]

$$CV = \frac{\sigma_v}{\langle v \rangle}, \quad \sigma_v = \sqrt{\langle v_i^2 \rangle - \langle v \rangle^2}. \tag{22}$$

To calculate PDF of the $x(t)$ time series for SFHN and EDFL models and $r_1(t)$ for RC, the length of 500 000 samples was always generated to obtain both true (x_t) and predicted (r_{1t}) signals.

IV. RESULTS

A. Strong and weak predictions in the SFHN model

We start our consideration with a model of a single stochastic neuron given by Eq. (11), under the influence of a noise source ξ with amplitude D, which was varied in the range (0.05, 1.0). Let us begin our consideration by numerically solving Eq. (11) using the EM numerical method of integration (4)—(5). Recall that, as

described in Sec. III B, we train the model at noise intensity $D = 0.2$ and optimize the hyperparameter values by simultaneously using a set of three noise amplitude values $D = \{0.05, 0.2, 1.0\}$. We consider the simplest situation where we have a single input noise signal, i.e., $M = 1$, and all information about the noise sources is undistorted $M_0 = M_e = 0$. The input vector of stochastic variables has $K = 2$ components and is formed as follows: $X_1 = x$, $X_2 = y$, where x and y are described by Eq. (11).

Figure 3 illustrates the results of prediction and their comparison with the true stochastic system signals (11) at three values of the noise amplitude: $D_1 = 0.2$, $D_2 = 0.6$, and $D_3 = 1.0$. Note that the RC training was performed at $D = D_1$, and each subsequent value of the noise amplitudes ($D_{2,3}$) introduces an increasing noise amplitude parameter mismatch of the true predicted system and the trained RC-based model. Figures 3(a), 3(d), and 3(g) show the time series of the predicted r and true x signals, and the dependence of the quality of the strong prediction γ (21) on time for the noise parameters D_1, D_2, and D_3, respectively.

An illustrative presentation demonstrating the presence (or proximity) of strong prediction is the dependence of the predicted signal r_{1t} on the target true x_t signal. The proximity of points to the diagonal $x_t = r_{1t}$ indicates that the RC has reached the regime of strong prediction of the stochastic system, while the widening of the diagonal and the appearance of a cloud of points indicate the destruction of strong prediction. Accordingly, Figs. 3(b), 3(e), and 3(h) illustrate the dependencies of the predicted signal r_1 on the true target signal x for noise parameters D_1, D_2, and D_3, respectively. To evaluate the weak prediction, we calculate the PDF of the ISI for the predicted r_{1t} and target x_t signals, which are shown in Figs. 3(c), 3(f), and 3(i) for different noise amplitudes.

From Figs. 3(a)–3(c), we can clearly see that the use of the RC allows us to realize the strong prediction mode when the SFHN system under study (11) and the trained RC [see Fig. 1(c)] are exposed to $D_1 = 0.2$. We observe a complete repetition of the stochastic system trajectory by the reservoir model. The prediction is strong ($\gamma(t) = 1$) throughout the RC model testing time. As might be expected, the PDFs of the ISI times are the same for the true and predicted stochastic processes. The coefficients of ISI variations (22) are also very close for the true and predicted processes: $CV_{\mathrm{FHN}} = 0.298$ and $CV_{\mathrm{RC}} = 0.294$.

In prediction behavior of the stochastic system at $D = 0.6$, we observe that the intervals of strong prediction characterized by high prediction quality ($\gamma = 1$) alternate with intervals of strong prediction that breaks down ($\gamma = 0$). Such a regime is illustrated in Fig. 3(d), where the switching of the prediction mode is clearly visible, which is reflected in the switching of the $\gamma(t)$ function from 1 to 0 and vice versa. The diagonal arrangement of points on the (x_t, r_{1t}) plane, characteristic of strong prediction, breaks down—a cloud of points appears, which shows that the trajectories of the true and predicted processes at certain moments of time no longer correspond to each other. At the same time, the shape of the PDFs of the ISIs of the true and predicted stochastic processes also remains almost the same [see Fig. 3(f)], as well as the CV values for the true and predicted processes: $CV_{\mathrm{FHN}} = 0.347$ and $CV_{\mathrm{RC}} = 0.348$. This corresponds to the weak prediction regime, but nevertheless, there are time intervals when the strong prediction regime is still observed. This mode of irregular switching between strong and weak prediction is of particular interest and will be discussed in Sec. IV B.

Finally, at $D = 1.0$, we do not observe a strong prediction regime. The prediction quality $\gamma(t) = 0$ during the whole analysis time [see Fig. 3(g)], and on the (x_t, r_{1t}) plane, we observe a cloud of points without a pronounced diagonal (Fig. 3h). However, the analyzed statistical characteristics of the stochastic processes remain the same [Fig. 3(i)], including almost the same coefficients of variation: $CV_{\mathrm{FHN}} = 0.392$ and $CV_{\mathrm{RC}} = 0.395$. According to our classification, this regime is a typical example of the weak prediction regime when the RC-based model does not allow us to predict the trajectory of a stochastic process even with a known stochastic influence on the initial system, but the model allows us to estimate with high accuracy the statistical characteristics of the predicted stochastic process.

Now, an interesting question arises: How does the control parameter, namely, the noise amplitude D affect the predictive ability of the model? It turns out that the quality of the model prediction greatly depends on the hyperparameter optimization strategy. We consider two strategies: (i) a standard strategy with hyperparameter selection only when testing the prediction on the same control parameter (noise amplitude $D_0 = 0.2$ at which the RC was trained) and (ii) a modified strategy for selecting optimal hyperparameters when testing on a set of three noise amplitudes ($D = \{0.05, 0.2, 1.0\}$).

In considering these issues, we will also compare the results of predicting the behavior of a stochastic neuron obtained by using different numerical methods for integrating stochastic equations—the EM method (4)—(5) and the EH method (6)—(8). Figure 4 illustrates the effect of coherence resonance in an SFHN neuron obtained by using EM and EH methods. As one can see, the only small difference is observed for $D = 0.05$ corresponding to a very low noise, which causes rare spikes generation. For all other noise's amplitudes, the CVs [see Eq. (22)] are almost the same and do not depend much on the integration method. This suggests that the integration method has little or no influence on the observed effects in the original stochastic system.

Next, we consider the effect of the numerical integration method of the original stochastic equations on the quality of the RC prediction. In this case, we use the time series obtained by integrating Eq. (11) using either the EM method or the EH method to train the reservoir.

Figure 5 shows the characteristics of the strong prediction quality—the parameter δ (20)—and the statistical characteristic estimated the weak prediction quality—the CV values and differences between predicted (CV_{RC}) and true (CV_{SFHN}) values,

$$\Delta CV = |CV_{\mathrm{RC}} - CV_{\mathrm{SFHN}}|. \tag{23}$$

Figure 6 illustrates the dependencies of the predicted signal r_{1t} on the true target x_t signal for different values of the noise amplitude D obtained with the help of an EM numerical method. The left column in these figures corresponds to the standard strategy, while the right column corresponds to the modified strategy of the RC-based model testing.

The former strategy is characterized by a very accurate strong prediction of the stochastic system behavior at the noise value $D \approx 0.2$ at which the RC was trained. As can be seen from Fig. 5(a) in the range $D \in (0.2, 0.3)$ for the EM method and $D \in (0.15, 0.3)$ for the EH method, we have a value $\delta > 0.95$

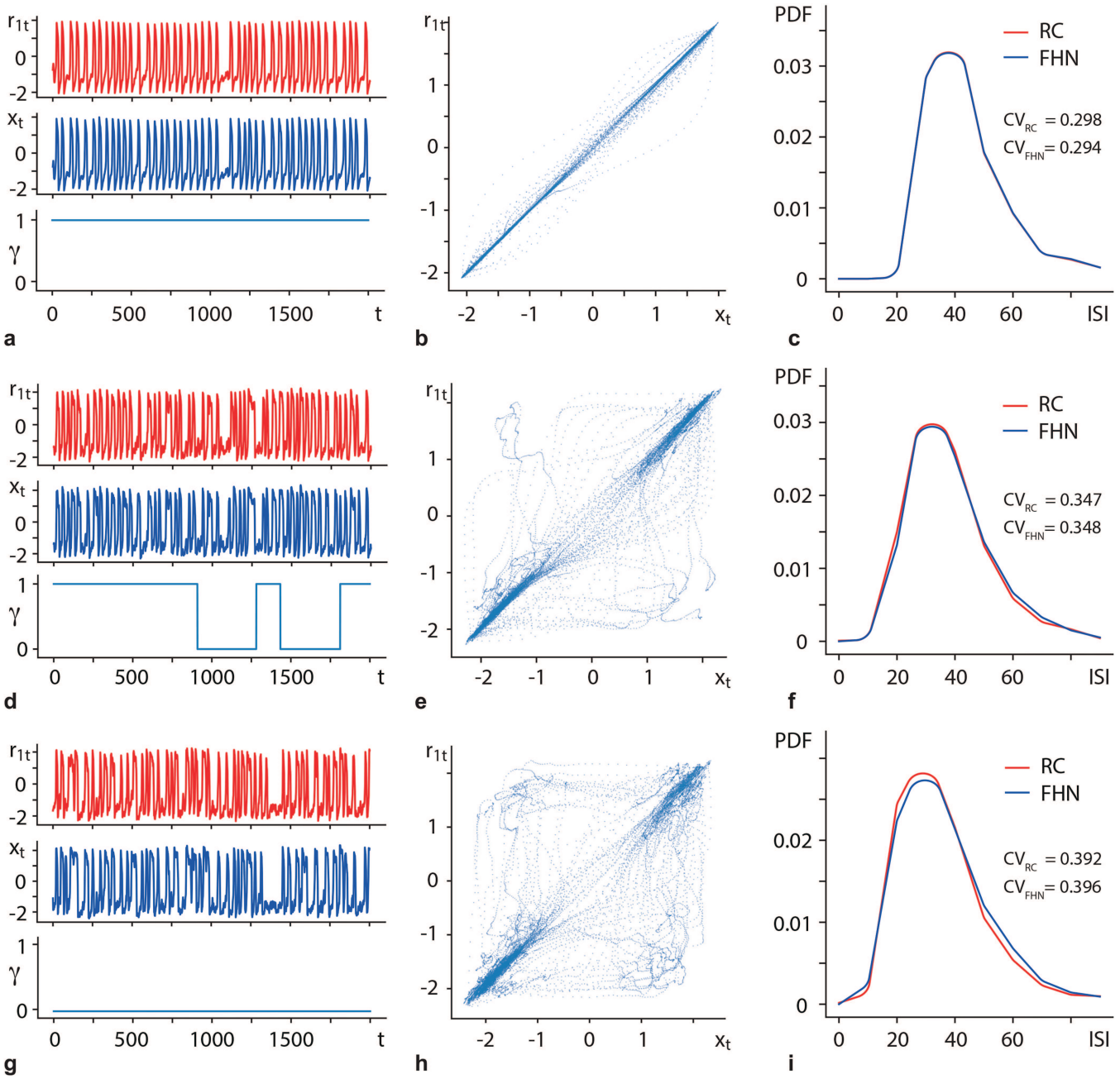

FIG. 3. Examples of strong and weak predictions of SFHN model dynamics. Panels (a), (d), and (g) show the time series of predicted r_{1t} (red) and true target x_t (blue) signals, as well as the quality dependence of the strong prediction on time γ. Panels (b), (e), and (h) show the dependence of predicted r_{1t} on the target x_t signal. Panels (c), (f), and (i) illustrate the PDF of ISI for the predicted and target signals. Noise amplitude values during the testing mode: (a)–(c) $D = 0.2$, (d)–(f) $D = 0.6$, and (g)–(i) $D = 1.0$.

(threshold at which $\gamma = 1$ is marked in the figure by a dashed line), which corresponds to a strong prediction. In the left panel of Fig. 6(b), we see a well-pronounced diagonal, which also indicates a strong prediction observation. However, when we start to vary the noise amplitudes D over a wide range of parameters, the region of weak prediction turns out to be quite small. This is illustrated by Figs. 5(c) and 5(e), which show the dependence of CV on the noise amplitude D in the original stochastic system and the predicted

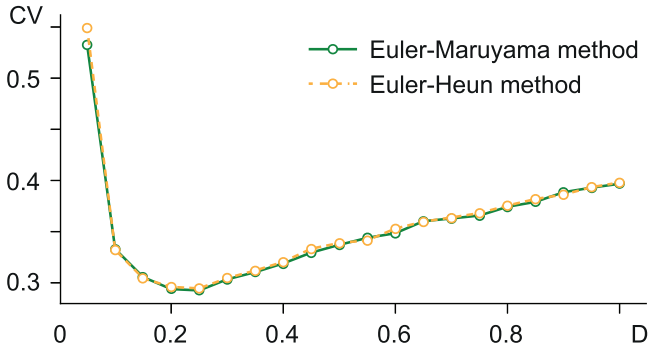

FIG. 4. Coherence resonance in an original stochastic neuron. Dependencies of the coefficients of variations [*CV*, Eq. (22)] of the SFHN neuron's signal obtained by using EM and EH methods of numerical integration of Eq. (11).

RC model. It can be seen that the differences between predicted and true values ΔCV appear small only at weak noise intensities $0.1 \leq D \leq 0.3$, while at noise intensities $D > 0.3$, the model stops predicting the behavior of the stochastic system even in a weak sense. From Fig. 5(c), it follows that the RC does not predict qualitatively and quantitatively the effect of coherent resonance in the model of a noise-excited neuron.

The modified RC testing strategy for hyperparameter selection proves to be more successful, as shown in Figs. 5(d) and 5(f). In this case, we have a weak prediction mode of the RC-based model over the whole analyzed range of the noise amplitude ($0.1 \geq D \geq 1$). Figure 5(d) shows that the coherent resonance curve is reproduced with a high degree of accuracy for both EM and EH numerical methods, and the difference between predicted and true values $\Delta CV < 0.02$ remains very small over the whole range of noise amplitudes. This is illustrated in Fig. 6 for large noise amplitudes ($D > 0.3$). On the left panels for the standard strategy, we have a complete collapse of the diagonal $x_t = r_{1t}$, while on the right panel, we observe typical behavior for the weak prediction regime. The strong prediction, as follows from Fig. 5(b), occurs at $0.25 \geq D \geq 0.4$ and at $D = 0.1$ for the EM method and at $0.2 \geq D \geq 0.4$ for the EH method. Interestingly, if we use RC trained on data obtained with the EM method, no strong prediction is observed at $D = 0.2$ [vertical line in Fig. 5(b)], the value of the noise intensity at which the reservoir was trained.

The obtained results show that the modified strategy for selecting optimal hyperparameters when testing on a set of three noise amplitudes is more effective in producing a model that better generalizes across different noise characteristics and provides strong and weak predictions over a wider range of the control parameter. However, there may be failures in strong prediction at the values of the control parameters on which the model was trained. At the same time, if there is a need for accurate prediction of the system at some fixed set of control parameters, the standard strategy of selecting hyperparameters of the RC model turns out to be more effective. We can also conclude that the obtained results do not depend much on the method of numerical integration of stochastic differential equations. Therefore, for simplicity of further analysis, we will use the data obtained with the EM method.

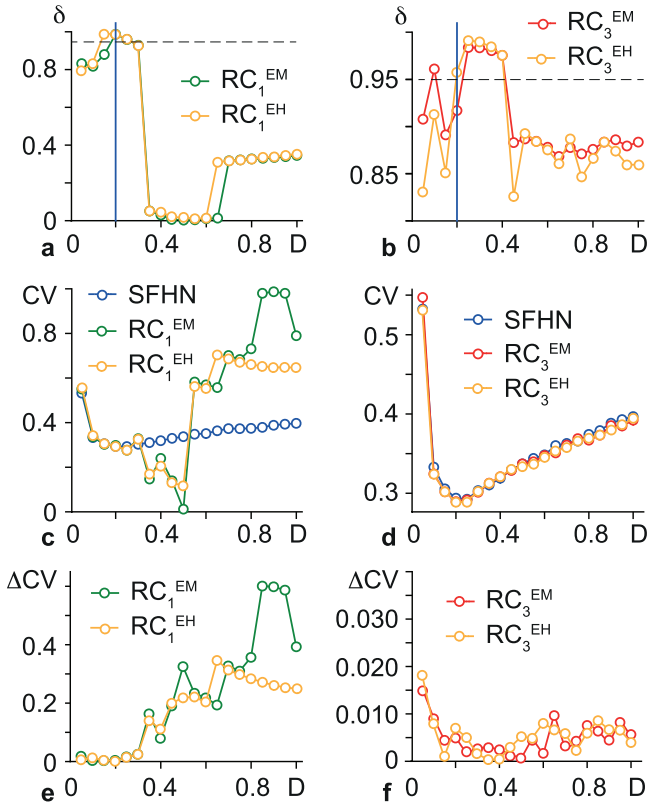

FIG. 5. Characteristics of stochastic neuron dynamics prediction vs the noise amplitude. Dependencies of (a) and (b) the characteristics δ of the strong prediction quality, (c) and (d) the CV values of predicted (CV_{RC}) and true SFHN (CV_{SFHN}) stochastic processes, and (e) and (f) the differences $\Delta CV = |CV_{RC} - CV_{SFHN}|$ on the noise amplitude D calculated by using EM and EH numerical methods. The left column with (a), (c), and (e) panels corresponds to optimization at the single value $D_0 = 0.2$ at which the reservoir was trained and the right column with (b), (d), and (f) panels—optimization simultaneously at three values of noise amplitude $D = \{0.05, 0.2, 1.0\}$. In panels (a) and (b), the dashed horizontal line corresponds to the threshold value $\delta_0 = 0.95$ [see Eq. (21)]; the vertical solid line corresponds to the noise value $D_0 = 0.2$ at which the RC was trained. True SFHN's CV on panels (c) and (d) was calculated by using the EM method [compare with Fig. 4].

B. Intermittency in strong prediction

Let us focus more on the irregular switching between strong and weak prediction modes, which was discussed in Sec. IV A. Indeed, beyond the boundary of strong prediction by noise amplitude at $D > 0.3$ in the stochastic neuron (11) [see Fig. 3(b)], we observe an intermittency regime that corresponds to switching between the strong and weak prediction regimes. This behavior is depicted in Fig. 7(a), which shows the dependence of $\gamma(t)$ over a long time interval at noise amplitude $D = 0.6$. The switches, whose duration L varies from event to event, are clearly visible in Fig. 7(a). Figure 7(b) illustrates the distribution of the durations of the strong prediction phases, which are plotted in a log–log scale for the same noise amplitude $D = 0.6$. A linear regression and a 95% confidence

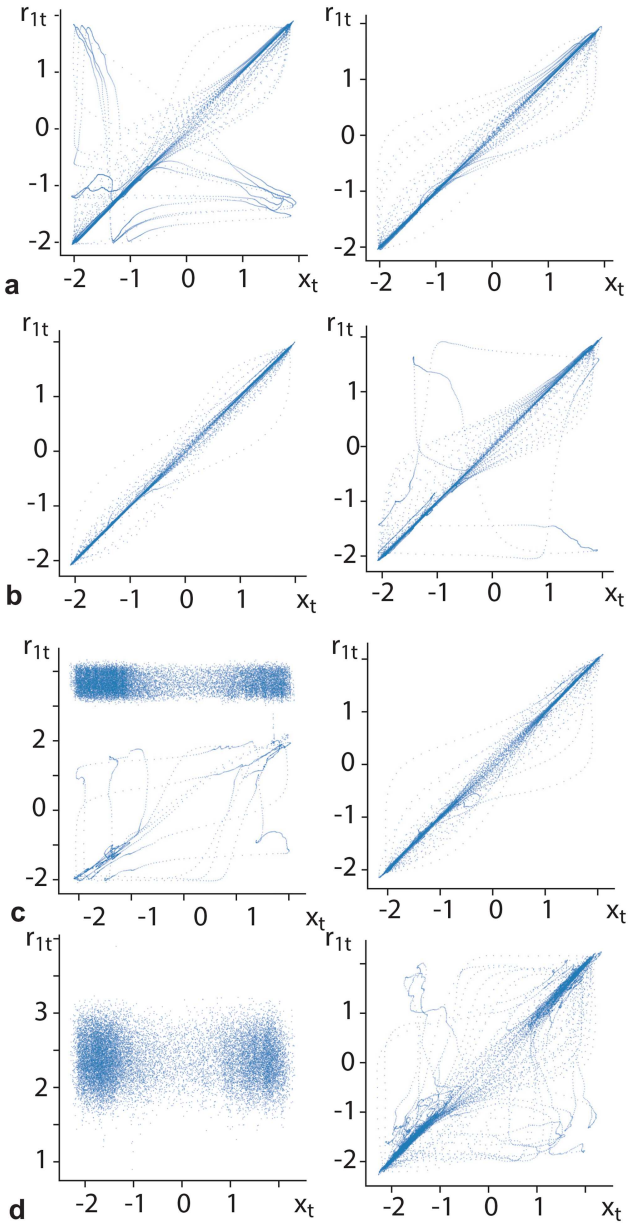

FIG. 6. Dependencies of the predicted r_{1t} signal on the true target x_t signal for different values of the noise D and the optimization strategy of the reservoir hyperparameters. The left column corresponds to optimization at the single value $D_0 = 0.2$ at which the reservoir was trained and the right column corresponds to optimization simultaneously at three values of noise amplitude $D = \{0.05, 0.2, 1.0\}$. The noise amplitude values are (a) $D = 0.1$, (b) $D = 0.2$, (c) $D = 0.4$, and (d) $D = 0.8$.

interval are also plotted as a solid line in Fig. 7(b). It is clearly seen that the dependence $N_L(L)$ is well approximated by the power law $N_L \propto L^{\eta}$, where $\eta \approx -1.5$. This dependence of the distribution of the number of phases N_L on their duration L resembles the property

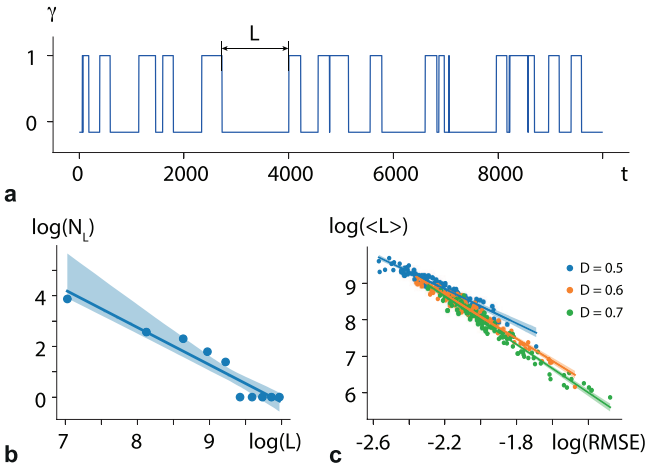

FIG. 7. Intermittent switching between strong and weak prediction phases. (a) Time dependence of the measure of accuracy of strong prediction $\gamma(t)$, illustrating the switching between phases at $D = 0.6$. (b) Corresponding distribution N_L of the durations L of the strong prediction phases ($\gamma(t) = 1$) plotted in a log–log scale. The solid line corresponds to a linear approximation with slope coefficient $\eta = -1.47$, and the blue region is a 95% confidence interval. (c) The dependencies of the mean duration $\langle L \rangle$ of the strong prediction phases on the RMSE plotted in a log–log scale for three noise amplitudes: $D = 0.5$, 0.6, and 0.7. Solid lines correspond to the linear approximations with coefficients (a) -2.35 (p-value $= 1.98 * 10^{-56}$), -3.04 (p-value $= 2.16 * 10^{-67}$), and -3.3 (p-value $= 2.25 * 10^{-70}$), respectively.

of on–off intermittency,[54,55] which, in particular, is observed at the boundaries of generalized synchronization[56,57] and noise-induced synchronization.[58]

If we consider the scheme shown in Fig. 2(c), which is implemented in the prediction of the trained RC model (see Fig. 2(b) with a training procedure illustration), we will see a great similarity to the auxiliary system method, which is implemented in the diagnosis of generalized synchronization.[59] The relationship between generalized synchronization and the construction of predictive models based on reservoir computing was also discussed in Ref. 60. Recall that generalized synchronization is observed in the unidirectionally coupled nonidentical self-sustained oscillators and assumes the presence of functional dependence between the drive $\mathbf{x}_d$ and the response $\mathbf{x}_r$ systems,

$$\mathbf{x}_r(t) = \mathbf{F}(\mathbf{x}_d(t)). \tag{24}$$

The functional relationship $\mathbf{F}(\cdot)$ is so complex that the diagnosing functional dependence is not a trivial task. Therefore, we usually use the auxiliary system method,[59] which involves the introduction of a replica $\mathbf{x}'_r$ of a response system, which is identical to the true response system but starts from other initial conditions. Due to the presence of functional coupling (24), after some transient, the regime is established when $\mathbf{x}_r(t) = \mathbf{x}'_r(t)$. It is shown in Ref. 61 that noise-induced synchronization is a special case of generalized synchronization when the influence of noise $\xi(t)$ dominates over the influence of dynamical system $\mathbf{x}_d$ itself. Coming back to Fig. 2(c), we see, in fact, the scheme of the auxiliary system approach, where the trained RC-based model acts as a replica of the response system. This

allows for considering the strong prediction mode, which consists of the coincidence of the trajectories of the predicted system and the reservoir $\mathbf{x}(t) = \mathbf{r}(t)$, as the mode of implementation of the auxiliary system method when generalized synchronization occurs. In our case, the coincidence of signals $\mathbf{x}(t) \approx \mathbf{r}(t)$ corresponds to the strong prediction mode, not to generalized synchronization, as in the case of truly unidirectionally coupled systems. In favor of this concept is also evidenced by Refs. 62 and 63 in which attempts were made to replace the auxiliary system with artificial neural networks. The possibility of predicting the behavior of the response system based on the signals of the drive system using machine learning means in fact establishing generalized synchronization.

In Ref. 57, it was shown that on–off intermittency occurs at the boundary of generalized synchronization, which is characterized by a power-law distribution of the duration of synchronous phases with degree exponent $\eta = -1.5$ and a power-law distribution of the average duration of the synchronization regime from the supercriticality parameter (usually the difference of coupling strengths $(\varepsilon - \varepsilon_c)$, where ε_c is the boundary of the onset of generalized/noise-induced synchronization) with the degree exponent $\beta = -1$. In our case, it is easy to check the distribution of durations of strong prediction temporal segments, and they are well described by a power law with exponent $\eta \approx -1.5$. However, it is not possible to estimate the degree of supercriticality because different RCs exhibit different prediction accuracy and, hence, different degrees of closeness to the mode of strong prediction. However, we can assume that the measure that characterizes the closeness of some RC prediction to the strong prediction mode is the RMSE value, which correlates with the prediction accuracy. In other words, the better the trained RC model, the smaller the RMSE and the lower the supercriticality.

Figure 7(c) shows the corresponding dependencies of the average duration $\langle L \rangle$ of the strong prediction phases on the RMSE for the three noise amplitudes tested. We used the following methodology to build the dependencies. We fixed the hyperparameter values, the average node degree $\bar{k} = 16$, and the spectral radius $\lambda = 0.6$ obtained during optimization at three values of the noise amplitude $(D = 0.05, 0.2, 1.0)$, and then 100 $\mathbf{H}_i$ $(i = 1, \ldots, 100)$ reservoir layer adjacency matrices with these hyperparameters were randomly generated. Each of the RCs described by the matrix $\mathbf{H}_i$ is trained on the signal at $D = 0.2$. Using the trained ith RC, predicted signals $r_{1t,i}$ of 500 000 samples duration were generated at different noise amplitudes D. For each signal, the average duration $\langle L \rangle_i$ of the strong prediction phases, and the corresponding RMSE_i values are calculated. The corresponding points $(\mathrm{RMSE}_i, \langle L \rangle_i)$ $(i = 1, \ldots, 100)$ are shown in Fig. 7(c) at $D = 0.5$, 0.6, and 0.7, respectively. For each D, a linear regression and a 95% confidence interval are also plotted in Fig. 7(c).

From Fig. 7(c), we see that the dependence of the average duration $\langle L \rangle$ of the strong prediction phase on *RMSE* is described by a power-law dependence with the degree exponent lying in the range $(-2.3, -3.3)$. The degree exponent decreases with increasing noise amplitude, i.e., as D is increased, the average duration of the strong prediction phase decreases with a simultaneous decrease in prediction accuracy estimated by *RMSE*. This is illustrated in Fig. 8, which shows the dependence of the average duration of the strong prediction phase and the value of *RMSE* on the noise amplitude D at

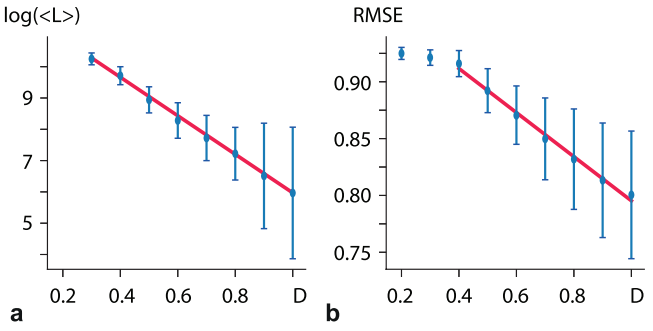

FIG. 8. Characteristics of intermittency and prediction quality under varying noise amplitudes. The dependencies of (a) the logarithm of the mean duration $\langle L \rangle$ of the strong prediction phase and (b) the mean value of *RMSE* on the noise amplitude D. The dots correspond to the mean values over all 100 RC under study and the whiskers correspond to the standard deviation. Red solid lines correspond to the linear approximation with coefficients (a) -6.16 (p-value $= 2.82 \times 10^{-9}$) and (b) -0.19 (p-value $= 1.25 \times 10^{-6}$).

which the RC was tested. The calculation is performed on all 100 RCs that were generated using the procedure described in the paragraph above. The dots in Fig. 8 show the mean values over all 100 reservoirs, and the whiskers mark the standard deviation. It can be clearly seen in Fig. 8(a) that the logarithm of the average duration of the strong prediction phase decreases linearly with increasing noise intensity, while the spread of this value also increases with increasing noise intensity, i.e., the amount of detuning of the noise amplitude D from the $D = 0.2$ value at which the RC was trained. The mean RMSE decreases linearly with increasing noise intensity at $D > 0.4$ [see Fig. 8(b)], that is, in the region where we observe the intermittency effect of strong prediction.

From Fig. 8(b), an important conclusion follows. In the region $D \leq 0.3$ of strong prediction, the variation of prediction accuracy when generating a random configuration of the reservoir layer network with optimally chosen hyperparameters is very small; i.e., by determining the value of hyperparameters, we can choose actually any variant of the reservoir layer network. However, the situation changes in the weak prediction regime. The specific topology of the reservoir layer network begins to have a strong influence on the prediction accuracy, so it is necessary to generate a certain number of networks at a given set of hyperparameters and select the best one from them in terms of prediction quality.

The on–off intermittency effect arises from the limited forecasting accuracy of the RC model. As an approximation of the original system, the RC model operates with minimal but non-zero prediction error. Both the original system and the RC model function in a sub-threshold regime, driven by a shared noise source. If the systems were identical ($\gamma = 0$), no forecasting errors or intermittency would occur. However, slight discrepancies between the systems lead to divergent behaviors under noise: some noise values result in negligible differences, while others cause significant trajectory deviations, and as consequence, the emergence of large difference between them ($\gamma = 1$). Due to the characteristic spike generation regime under low noise, both systems stabilize into a stationary state, corresponding to accurate predictions. This laminar

phase persists until noise-induced divergence disrupts the identical behavior, terminating the accurate prediction phase. The interplay of these mechanisms generates the observed on–off intermittency in prediction accuracy.

C. Two coupled stochastic neurons

Let us now consider two coupled stochastic neurons described by Eq. (12), fed with independent Gaussian noises: $D_1\xi_1$ and $D_2\xi_2$. For definiteness, the noise amplitudes are chosen to be the same for each neuron ($D_1 = D_2 = D$). When considering this problem, we are primarily interested in the influence of the completeness of the description of noise sources on the prediction quality of the stochastic system dynamics. Recall that, as we discussed in Sec. III B 1, the input vector $\mathbf{g}$ (14) generally includes $(M - M_0 + M_e)$ noise components, where M is the true value of noise sources that affect the system dynamics (in our case, $M = 2$), M_0 is the number of noise sources about which we have no information but which influence the behavior of the stochastic system, and M_e is the number of sources that we mistakenly take into account to describe the dynamics of the stochastic system while they do not affect its behavior.

Therefore, we consider three situations that illustrate to some extent the above situations of incomplete data about noise sources on a stochastic system. The first situation corresponds to the case when we train the RC on complete information; i.e., the following input vector (14) is used in training:

$$\mathbf{g}_t = \left(D\xi_{1,t}, D\xi_{2,t}, x_{1,t}, x_{2,t}, y_{1,t}, y_{2,t},\right)^T. \tag{25}$$

In other words, for the considered system (12), we have $M = 2$, and we know all noise sources; i.e., $M_0 = M_e = 0$.

In the second situation, we consider the case where we do not know all information about the noise sources; namely, we do not have information about one of the noise effects (for certainty $D\xi_{2,t}$), that is, $M_0 = 1$ and $M_e = 0$. In this case, the input vector has the form

$$\mathbf{g}_t = \left(D\xi_{1,t}, x_{1,t}, x_{2,t}, y_{1,t}, y_{2,t},\right)^T. \tag{26}$$

Finally, we consider the third situation when in addition to the situation of incomplete input information, we mistakenly consider an additional noise source $D\xi_e$ with the same statistical characteristics as the first noise source, but not affecting the dynamics of the original system (12). That is, the input signal in this case is described by two noise sources, but at the same time, $M_0 = 1$ and also $M_e = 1$:

$$\mathbf{g}_t = \left(D\xi_{1,t}, D\xi_{e,t}, x_{1,t}, x_{2,t}, y_{1,t}, y_{2,t},\right)^T. \tag{27}$$

For all three situations, we use the same metrics to analyze the quality of prediction for each of the coupled neurons.

To determine the modes of strong and weak predictions, we use $\gamma_{1,2}$ [see Eq. (21)] and ΔCV [Eq. (23)] features, which are calculated for each of the two neurons separately. A strong prediction corresponds to $\gamma_{1,2} = 1$, while a weak prediction for each of the neurons to $\Delta CV_{1,2} < 0.075$.

We investigate the parameter space "noise amplitude D—coupling strength between neurons ρ." The corresponding results are illustrated in Fig. 9, which shows the parameter spaces for all three situations. Each panel shows four main modes of an RC prediction: (i) a strong prediction of the dynamics of both stochastic neurons (white area), (ii) a weak prediction of the dynamics of both stochastic neurons (green area), (iii) a weak prediction of the dynamics of only one of the neurons (blue area), and (iv) no prediction (black area).

Given the dependence of the prediction accuracy on the particular reservoir's adjancency matrix $\mathbf{H}$, we calculate the averaged parameter space (D, ρ). To do this, the following procedure is carried out. We start by generating a reservoir's adjancency matrix $\mathbf{H}_s$. We then optimize the hyperparameters for this reservoir layer by enumerating the values of the spectral radius and the average node degree for the data obtained at $D = 2$ and selecting the most optimal pair of hyperparameters at three noise amplitudes $D = \{0.05, 0.2, 1.0\}$ in terms of RMSE, as discussed in Sec. III B 4. This procedure is repeated for all values of coupling strength $\rho_i = i\Delta\rho$ $(i = 1, \ldots, 10,\ \Delta\rho = 0.1)$. Then, for selected values of the hyperparameters for each value of ρ_i, stochastic process predictions for the entire range of values of the noise amplitude D are carried out for a fixed value of the coupling strength ρ_i. As a result, we obtain a particular parameter space (ρ, D) for which the particular values of $\gamma_s(\rho, D)$ and $\Delta CV_s(\rho, D)$ are known. This procedure is repeated $S = 10$ times, with the specific values of the hyperparameters differing in each calculation due to the randomness of the generated reservoir's adjancency matrix. We then average the particular values of the prediction quality characteristics over the S optimizations and obtained the average values of $\langle\gamma_{1,2}(D, \rho)\rangle$ and $\langle\Delta CV_{1,2}(D, \rho)\rangle$, from which the averaged parameter space is reconstructed.

The resulting averaged parameter spaces "noise amplitude D—coupling strength between neurons ρ"—are shown in Figs. 9(a)–9(c) for all three situations of input vector formation (25)–(27), respectively. The comparison of the parameter spaces for the three situations shows that the best prediction quality occurs when we train the RC-based model on the full information set of a stochastic system with multiple noise sources. It can be seen from Fig. 9(a) that we have a strong prediction at all coupling strengths $\rho \in [0, 1]$ at noise amplitude $D = 0.2$ at which the RC was trained. We also have in almost the entire remaining parameter space a weak prediction of the behavior of two coupled stochastic neurons. This is an important result that shows that a RC-based model trained at one value of a noise amplitude allows for predicting the statistical characteristics of the dynamics of coupled stochastic neurons when the noise amplitude is varied.

However, a strong prediction, as in the case of a single stochastic neuron, at noise amplitudes other than $D = 0.2$, at which the RC was trained, cannot be achieved. This is probably due to the more complex nature of the excitation of spikes in each coupled neuron, which is now determined not only by noise but also by dynamics of the coupled neurons, making strong prediction for the RC impossible. This is indirectly confirmed by the fact that at strong noise with $D > 0.6$, when spikes are frequently generated and their sequence is characterized by high $CV_{1,2}$, at large coupling coefficients $\rho > 0.8$, the model loses the ability to provide even weak prediction of the statistical characteristics of both neurons. In this case, we deal with either a weak prediction of the stochastic dynamics of only one neuron or no any prediction.

The other two situations with insufficient or erroneous information about the stochastic system that was used to train the

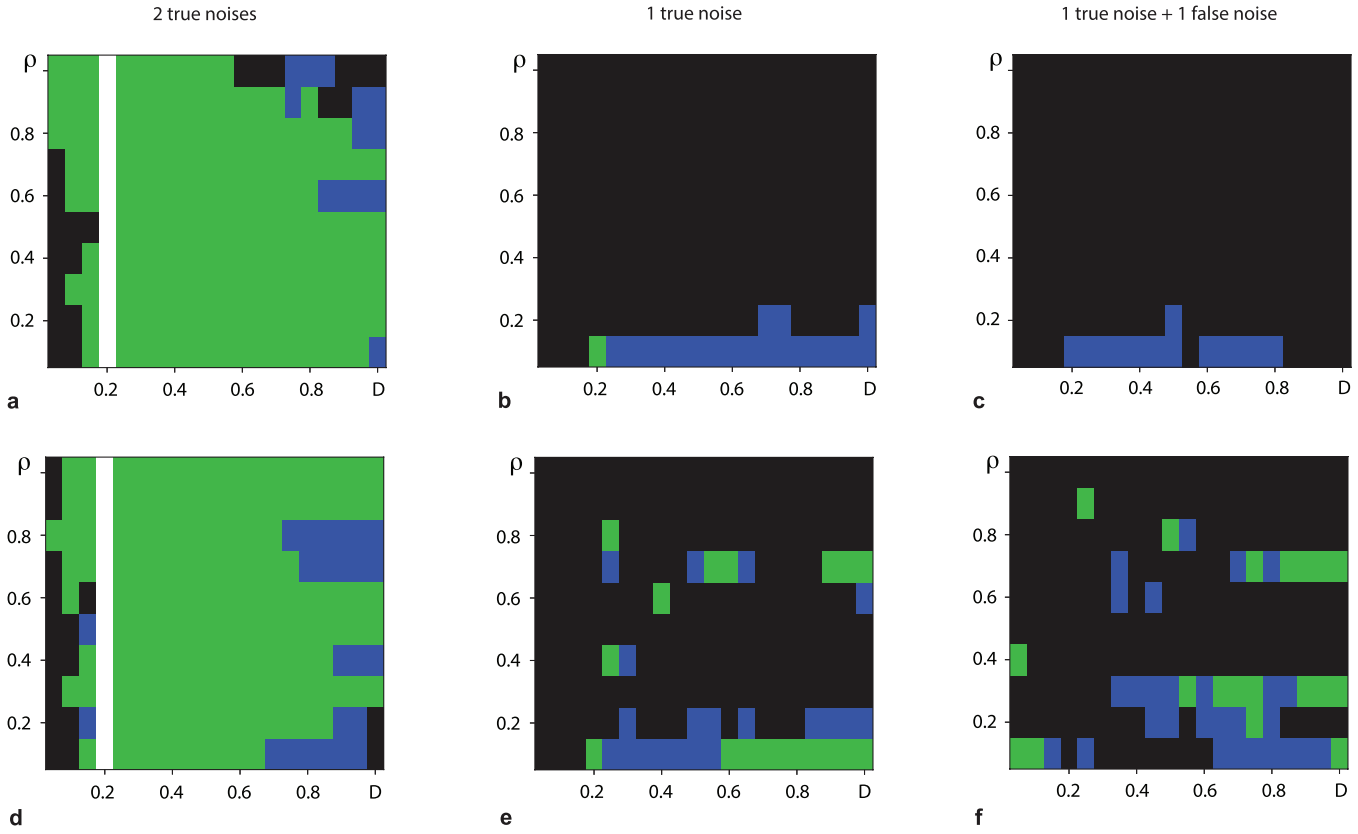

FIG. 9. Prediction mode areas on the parameter spaces "noise amplitude D—coupling strength ρ" for a different situation of input vector formation. Averaged (top) and obtained with random adjacency matrix $\mathbf{H}_s$ (bottom) parameter space "noise amplitude D—coupling strength ρ between neurons" for a system consisting of two coupled neurons for three model situations: (a) and (d) two noises, each of which corresponds to real noise, are fed to the input of the RC, fed to each neuron; (b) and (e) only one noise, which corresponds to real noise, is fed to the input of the reservoir, fed to the first neuron; (c) and (f) two noises are fed to the input of the RC, one of which corresponds to real noise, fed to the first neuron, and the other is a separately generated noise with the same statistical characteristics but absent in the original system. Here, white color defines an accurate prediction, green color a weak prediction of both neurons, blue color a weak prediction of only one neuron, and black color absence of strong and weak predictions of both neurons.

RC show much worse prediction capabilities. In the second situation, as can be seen in Fig. 9(b), we have the possibility of a weak prediction only at a small coupling coefficient $\rho = 0.1$ and at $D = 0.2$ (the noise amplitude at which the reservoir was trained). In this situation, when the neurons practically do not influence each other, it is enough to feed a single noise into the system to force both neurons exhibit dynamics whose statistical characteristics would correspond to the original stochastic system. At $\rho = 0.1$ and other noise values with $D > 0.2$, we have a situation when we can predict the statistical characteristics of the dynamics of only one of the neurons. For larger values of the coupling coefficient, we cannot predict the behavior of the stochastic system even in a weak sense. In the third situation, when we add additional error noise ξ_e (27), the RC is unable to give even a weak prediction [see Fig. 9(c)]. At small coupling coefficients in the noise range $D \in (0.2, 0.8)$, we can only predict statistical characteristics of spike generation at the first neuron whose noise exposure is known to the RC.

It should be noted that the conclusions drawn are valid for the averaged parameter spaces over S trained RCs with randomly generated adjacency matrices. If we consider particular sth variants of RC with some particular random adjacency matrix $\mathbf{H}_s$, then local improvements are possible for the second and third situations considered. As an illustration, Figs. 9(d)–9(f) show the particular parameter spaces obtained by considering a particular RC with matrix $\mathbf{H}_s$. It is clearly seen that in these situations of input vector formation [see Eqs. (26) and (27)], it is possible to obtain a weak prediction for some combinations of parameters D and ρ even in the absence of all available noise source information [see green regions in Fig. 9(e)] and with erroneous noise information [see Fig. 9(f)].

The above discussed results suggest that, as already mentioned in Refs. 12, 60 and 64, the problem of selecting a specific reservoir adjacency matrix is an unsolved scientific problem, but the effectiveness of a particular RC-based model depends on it. This is largely determined by the specificity of the RC, which involves training the coupling weights only in the output layer, but not in the reservoir

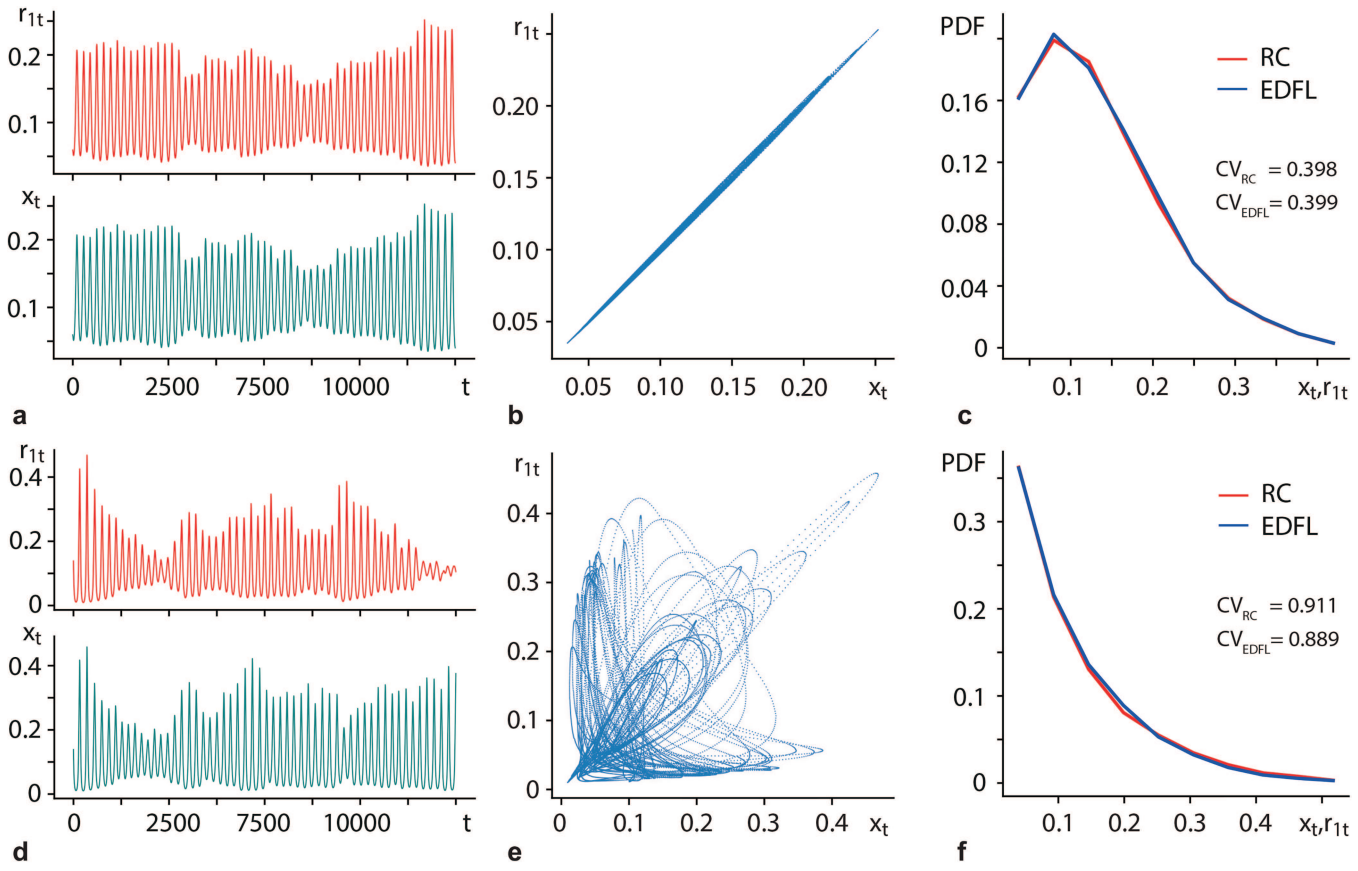

FIG. 10. Examples of strong and weak predictions of EDFL dynamics. (a) and (d) Time series of predicted r_{1t} (red) and true target x_t (green) signals. (b) and (e) Dependence of predicted r_{1t} on the target x_t signal. (c) and (f) PDF of the laser power for the predicted r_1 (blue) and target x (red) signals. The values of the noise amplitude during testing mode are (a) and (c) $D = 0.3$ and (d) and (f) $D = 0.7$. The RC was trained at noise intensity $D_0 = 0.4$.

layer. As a consequence, it is the specific adjacency matrix of the reservoir layer that is the factor whose enumeration will allow for an improvement in the quality of the model. At the same time, it should be kept in mind that the quality of the RC-based model is improved in an extensive way when a large number of reservoirs have to be generated and analyzed to obtain the desired result.

D. Strong and weak predictions in a noise-pumped erbium-doped fiber laser

Let us now consider the application of the considered approach to the EDFL model. Consider a reservoir learning situation where we have a single true noise source ($M = 1$ and $M_0 = M_e = 0$). The input vector of stochastic variables has $K = 2$ components and is formed as follows: $X_1 = x$, $X_2 = y$, where the variables x and y are described by Eq. (13).

Figure 10 illustrates the abilities of the reservoir trained and optimized at only one noise value $D = 0.4$ to strongly and weakly predict stochastic laser dynamics when varying the intensity of noise applied to the trained reservoir and, respectively, the SDEs (13). The hyperparameters of the optimal reservoir were as follows: average node degree $\langle k \rangle = 13$, spectral radius $\lambda = 0.3$. Figures 10(a)–10(c) obtained at noise $D = 0.3$ illustrate the effect of the strong prediction of the laser dynamics. It should be noted that RC training and prediction takes place at different noise intensities ($D = 0.4$ and $D = 0.3$, respectively). We can clearly see that a diagonal line is observed on the (x_t, r_{1t}) plane, which corresponds to the fulfillment of the condition (9) of a strong prediction mode and means that the true and predicted trajectories are practically identical: $x_t \approx r_{1t}$. Obviously, the distribution functions and the corresponding CV measures also turn out to be the same ($\Delta CV = |CV_{\mathrm{RC}} - CV_{\mathrm{EDFL}}| = 0.001$), as can be seen in Fig. 10(c).

Increasing variation of the noise intensity with respect to the noise value at which the RC was trained ($D = 0.4$) leads to the breakdown of the strong prediction mode, as illustrated in Figs. 10(d) and 10(e) corresponding to the driven noise intensity $D = 0.7$. It can be clearly seen that the time series of the RC and the true SDEs differ, and the diagonal on the (x_t, r_{1t}) plane collapses, indicating that the strong prediction mode cannot be realized. However, if we analyze the laser power amplitude PDF at a given noise intensity $D = 0.7$,

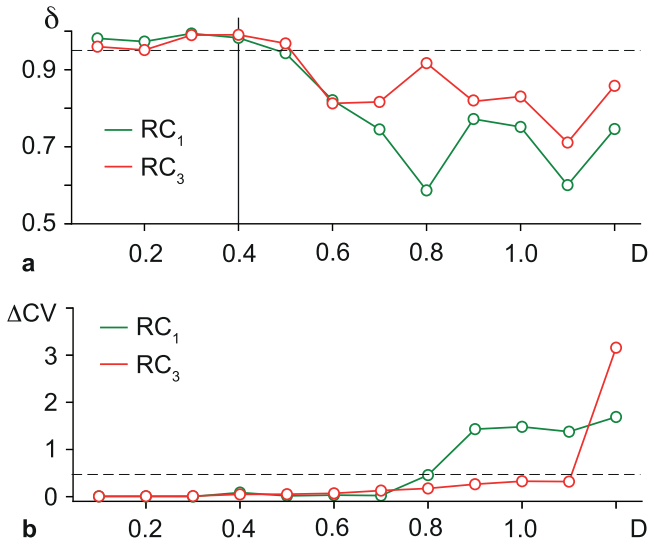

FIG. 11. Characteristics of EDFL dynamics prediction vs the noise intensity. (a) Characteristics δ (20) of strong prediction quality and (b) differences $\Delta CV = |CV_{RC} - CV_{EDFL}|$ vs noise amplitude D. The green lines correspond to the optimization at the single value $D_0 = 0.4$ at which the RC was trained. The red lines correspond to the optimization simultaneously at three values of the noise amplitude $D = \{0.1, 0.4, 1.2\}$. In panel (a), the dashed horizontal line corresponds to the threshold value $\delta_0 = 0.95$ [see Eq. (21), which is identical to the one introduced earlier for the stochastic neuron]; the vertical solid line corresponds to the noise value $D_0 = 0.4$ at which the RC was trained. In panel (b), the dashed horizontal line corresponds to the threshold value $\Delta CV = 0.5$ that can be considered a boundary of weak prediction.

we see that these PDFs almost coincide. This refers to the weak prediction mode in the RC at this noise intensity. To evaluate the weak prediction, we calculate the coefficients of variations (22) of the target laser x_t and predicted RC r_{1t} signals. The CV values were also very close for the true and predicted processes: $CV_{EDFL} = 0.899$ and $CV_{RC} = 0.911$ ($\Delta CV = 0.012$).

To analyze the possibility of extending the parameter region in which strong and weak prediction modes are possible, we compare two hyperparameter optimization strategies, as we did for the stochastic neuron (compare with Fig. 5 for the SFHN model). As before, we consider (i) a standard strategy with hyperparameter selection only when testing the prediction on the same control parameter (noise amplitude $D_0 = 0.4$ at which the RC was trained) and (ii) a modified strategy for selecting optimal hyperparameters when testing on a set of three noise amplitudes $D = \{0.1, 0.4, 1.2\}$.

Figure 11 shows the characteristics δ (20) of the strong prediction quality and the statistical characteristic estimated the quality of weak prediction, i.e., the differences ΔCV between predicted (CV_{RC}) and true (CV_{EDFL}) values.

The standard and modified strategies are characterized by a very accurate strong prediction of the behavior of the stochastic system at the noise intensities $D \in [0.1, 0.5]$. As can be seen from Fig. 11(a), in this range of D, we have a value $\delta > 0.95$ (threshold at which $\gamma = 1$ is marked in the figure by a dashed line), which corresponds to a strong prediction for both hyperparameter selection strategies. So, for the considered EDFL system, the hyperparameter selection strategy does not allow us to extend the region of a strong prediction in the parameter space.

At the same time, the region of a weak prediction expands significantly when using the modified strategy. In Fig. 11(b), we can clearly see that the difference ΔCV between predicted and true values for the modified strategy (red line) slowly grows up to the noise value $D = 1.1$ while remaining limited to $\Delta CV < 0.5$ (marked in the figure by the dashed line). For the standard strategy, we have a collapse of the weak prediction mode at $D > 0.8$.

Thus, as in the previous example of a stochastic neuron, a modified strategy for selecting optimal hyperparameters when testing on a set of three noise amplitudes is more effective in producing a model that better generalizes across different noise characteristics and provides weak predictions over a wider range of control parameters. However, unlike the case of the stochastic neuron, we cannot expand the strong prediction region, but we significantly expand the weak prediction region of the RC-based model.

V. CONCLUSION

In this paper, we suggested two different modes of predicting dynamics of stochastic systems using reservoir computing techniques. We proposed an approach to replicate a stochastic system and forecast its dynamics using reservoir computing driven by noise sources. We have shown that such RC-based models can generalize the behavior of the stochastic system within a wide range of control parameters. However, the quality of the prediction strictly depends on how we train RC. We introduced two modes for predicting the behavior of a stochastic system. When the testing parameters of the system are close to the training ones, we can achieve a strong prediction mode characterized by an almost true replica with a given accuracy of the stochastic system's trajectory determined by the noise influence and initial conditions. By detuning the parameters, we can only predict the characteristics of the probability distribution of the stochastic process, which corresponds to a weak prediction. We tested our approach on the model of a single stochastic FitzHugh–Nagumo neurons and showed that it was effective and allowed, within a wide range of noise parameters, to predict the characteristics of a stochastic system using a RC. Moreover, we found an intermittency effect when the model exhibited stages of a strong prediction that were followed by modes of a weak prediction. This behavior had some features of on–off intermittency, which was related to the features of establishing a generalized synchronization that was close to the strong prediction regime establishment condition between the predicted stochastic system and the trained reservoir.

We also analyzed a model of two coupled stochastic neurons, where we were primarily interested in the effect on the prediction of the completeness of the information on which the RC was trained. The lack of information led to a significant reduction in the prediction quality, which can only be made in a weak sense in the region of control noise parameters close to those at which the RC-based model was trained. However, there are some variants of the reservoir layer in which the prediction quality can be increased. This issue requires further consideration and research.

In addition, we considered the possibility of using our approach to analyze a physical system, the model of noise-pumped EDFL. We have shown that the proposed approach to design reservoir computing for strong and weak predictions of such a system works perfectly, allowing us to analyze and predict statistical characteristics of the stochastic laser system in a wide range of parameters. This allowed us to reach a conclusion about the universality of the proposed concept and its applicability to a wide range of excitable systems.

Finally, we addressed the question on prediction accuracy of stochastic systems we find in real life (financial markets, geophysical and climatic systems, control systems, etc.) and that are constantly affected by uncontrolled noise. In the considered scheme for a strong prediction, we need to collect the noise source data, which are not always possible, while in the case of weak synchronization, we do not need to know all the information about the noise, but only its statistical characteristics, and we can only predict the statistical characteristics of the process, but not its specific trajectory. For many problems, this is not enough, so it is important for further research in this direction to improve the quality of prediction models for stochastic processes and systems.

Our findings provide a powerful framework for solving real-world problems in neuroscience, laser physics, intelligent systems for autonomous devices, and other fields. Using these results, it is possible to design more efficient control systems, improve prediction accuracy, and exploit the role of noise in controlling complex behavior. For example, a strong or weak prediction of brain activity is critical for preventing various undesirable effects and diseases, as well as for applications in brain–machine interfaces.[65] A weak prediction, in particular, can help identify modes of neural activity or build models to study how noise and system parameters contribute to pathological states.[27,66,67] In intelligent systems of brain–computer interfaces, it is possible to use models with a weak prediction to estimate and predict the characteristics of noise in brain activity signals and accordingly deal with it effectively to improve the accuracy of classification of brain activity patterns, and models with a strong prediction even at small time intervals will allow one to determine these activity patterns more quickly and consequently increase the speed of information transfer in neurointerfaces.[68]

Although this study demonstrates the potential of RC to predict stochastic system dynamics, a number of limitations should be recognized. First, the current approach assumes that the stochastic dynamics is stationary at the prediction horizon. However, many real systems exhibit non-stationary behavior, such as time-varying parameters or abrupt regime changes. The RC model may fail to adapt to such changes without additional online learning or adaptation mechanisms. This calls for further research on these issues as applied to the analysis of systems exhibiting non-stationary behavior. Second, the proposed RC framework performs well on the tested systems, but its scalability for stochastic systems of high dimensionality remains uncertain. As the system dimensionality increases, the reservoir layer may require significantly larger sizes, leading to increased computational complexity and memory requirements. In particular, it raises the question of the predictability of stochastic systems described by partial derivatives where the noise source is distributed in space. Third, despite the empirical success of the RC in this study, the theoretical basis for its effectiveness for predicting stochastic systems remains incomplete. A deeper theoretical framework is needed to explain why and how RC captures the underlying dynamics of stochastic systems. Finally, the results presented in this study are based on specific stochastic systems. The generalizability of the approach to a broader class of stochastic systems, such as systems with large-tailed distributions or long-range correlations, remains to be confirmed. Further studies on a variety of systems are needed to establish the broader applicability of the proposed method. In particular, it is interesting to apply the approach to stochastic systems exhibiting intrinsic dynamics. This will also be the subject of further research.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation (Grant No. 23-71-30010).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Alexander E. Hramov: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal). **Nikita Kulagin:** Investigation (equal); Software (equal); Validation (equal); Writing – review & editing (equal). **Alexander N. Pisarchik:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Writing – review & editing (equal). **Andrey V. Andreev:** Conceptualization (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] I. Prigogine, "Time, structure, and fluctuations," Science **201**, 777–785 (1978).
[2] D. Kondepudi, T. Petrosky, and J. A. Pojman, "Dissipative structures and irreversibility in nature: Celebrating 100th birth anniversary of Ilya Prigogine (1917–2003)," Chaos **27**, 104501 (2017).
[3] A. Goldbeter, "Dissipative structures in biological systems: Bistability, oscillations, spatial patterns and waves," Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **376**, 20170376 (2018).
[4] T. Misteli, "The self-organizing genome: Principles of genome architecture and function," Cell **183**, 28–45 (2020).
[5] D. Witthaut, F. Hellmann, J. Kurths, S. Kettemann, H. Meyer-Ortmanns, and M. Timme, "Collective nonlinear dynamics and self-organization in decentralized power grids," Rev. Mod. Phys. **94**, 015005 (2022).
[6] A. Ghadami and B. I. Epureanu, "Data-driven prediction in dynamical systems: Recent developments," Philos. Trans. R. Soc. A **380**, 20210213 (2022).
[7] W.-X. Wang, Y.-C. Lai, and C. Grebogi, "Data based identification and prediction of nonlinear and complex dynamical systems," Phys. Rep. **644**, 1–76 (2016).

[8]A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," BMC Med Inform Decis Mak **19**, 211 (2019).
[9]Y. Sun, F. Haghighat, and B. C. Fung, "A review of the-state-of-the-art in data-driven approaches for building energy prediction," Energy Build. **221**, 110022 (2020).
[10]K. Nakajima and I. Fischer, *Reservoir Computing* (Springer, 2021).
[11]J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," Phys. Rev. Lett. **120**, 024102 (2018).
[12]D. J. Gauthier, E. Bollt, A. Griffith, and W. A. Barbosa, "Next generation reservoir computing," Nat. Commun. **12**, 5564 (2021).
[13]J. Z. Kim and D. S. Bassett, "A neural machine code and programming framework for the reservoir computer," Nat. Mach. Intell. **5**, 622–630 (2023).
[14]Z. Li, A. Andreev, A. Hramov, O. Blyuss, and A. Zaikin, "Novel efficient reservoir computing methodologies for regular and irregular time series classification," Nonlinear Dyn. **113**, 4045–4042 (2024).
[15]S. Shahi, F. H. Fenton, and E. M. Cherry, "Prediction of chaotic time series using recurrent neural networks and reservoir computing techniques: A comparative study," Mach. Learn. Appl. **8**, 100300 (2022).
[16]F. Köster, D. Patel, A. Wikner, L. Jaurigue, and K. Lüdge, "Data-informed reservoir computing for efficient time-series prediction," Chaos **33**, 073109 (2023).
[17]J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, "Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data," Chaos **27**, 121102 (2017).
[18]A. R. Ismail, S. Jovanovic, S. Petit-Watelot, and H. Rabah, "Detection of chaos using reservoir computing approach," IEEE Access **10**, 52686–52699 (2022).
[19]S. Saha, A. Mishra, S. Ghosh, S. K. Dana, and C. Hens, "Predicting bursting in a complete graph of mixed population through reservoir computing," Phys. Rev. Res. **2**, 033338 (2020).
[20]M. Roy, A. Senapati, S. Poria, A. Mishra, and C. Hens, "Role of assortativity in predicting burst synchronization using echo state network," Phys. Rev. E **105**, 064205 (2022).
[21]S. Pandey and J. Schumacher, "Reservoir computing model of two-dimensional turbulent convection," Phys. Rev. Fluids **5**, 113506 (2020).
[22]J. Pathak and E. Ott, "Reservoir computing for forecasting large spatiotemporal dynamical systems," in *Reservoir Computing: Theory, Physical Implementations, and Applications* (Springer, 2021), pp. 117–138.
[23]A. V. Andreev, A. A. Badarin, V. A. Maximenko, and A. E. Hramov, "Forecasting macroscopic dynamics in adaptive Kuramoto network using reservoir computing," Chaos **32**, 103126 (2022).
[24]M. Roy, S. Mandal, C. Hens, A. Prasad, N. Kuznetsov, and M. Dev Shrimali, "Model-free prediction of multistability using echo state network," Chaos **32**, 101104 (2022).
[25]B. Lindner, J. Garcıa-Ojalvo, A. Neiman, and L. Schimansky-Geier, "Effects of noise in excitable systems," Phys. Rep. **392**, 321–424 (2004).
[26]J. Wang, "Landscape and flux theory of non-equilibrium dynamical systems with application to biology," Adv. Phys. **64**, 1–137 (2015).
[27]A. N. Pisarchik and A. Hramov, "Stochastic processes in the brain's neural network and their impact on perception and decision-making," Usp. Fiz. Nauk **193**, 1298–1324 (2023).
[28]T. Wellens, V. Shatokhin, and A. Buchleitner, "Stochastic resonance," Rep. Prog. Phys. **67**, 45 (2003).
[29]V. S. Anishchenko, A. B. Neiman, F. Moss, and L. Shimansky-Geier, "Stochastic resonance: Noise-enhanced order," Phys.-Usp. **42**, 7 (1999).
[30]A. N. Pisarchik and A. E. Hramov, "Coherence resonance in neural networks: Theory and experiments," Phys. Rep. **1000**, 1–57 (2023).
[31]S. Aumaître, K. Mallick, and F. Pétrélis, "Noise-induced bifurcations, multiscaling and on–off intermittency," J. Stat. Mech.: Theory Exp. **2007**, P07016.
[32]A. Campos-Mejía, A. N. Pisarchik, and D. A. Arroyo-Almanza, "Noise-induced on–off intermittency in mutually coupled semiconductor lasers," Chaos, Solitons Fractals **54**, 96–100 (2013).
[33]M. Pradas, D. Tseluiko, S. Kalliadasis, D. T. Papageorgiou, and G. A. Pavliotis, "Noise induced state transitions, intermittency, and universality in the noisy Kuramoto-Sivashinksy equation," Phys. Rev. Lett. **106**, 060602 (2011).
[34]C. Van den Broeck, J. Parrondo, and R. Toral, "Noise-induced nonequilibrium phase transition," Phys. Rev. Lett. **73**, 3395 (1994).
[35]M. Ghodrati, "Complexity growth rate during phase transitions," Phys. Rev. D **98**, 106011 (2018).
[36]R. N. Mantegna and B. Spagnolo, "Noise enhanced stability in an unstable system," Phys. Rev. Lett. **76**, 563 (1996).
[37]K. Sobczyk, *Stochastic Differential Equations: With Applications to Physics and Engineering* (Springer Science & Business Media, 2013), Vol. 40.
[38]L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega, "Stochastic nonlinear time series forecasting using time-delay reservoir computers: Performance and universality," Neural Netw. **55**, 59–71 (2014).
[39]C. Fang, Y. Lu, T. Gao, and J. Duan, "Reservoir computing with error correction: Long-term behaviors of stochastic dynamical systems," Physica D **456**, 133919 (2023).
[40]Z. Liao, Z. Wang, H. Yamahara, and H. Tabata, "Low-power-consumption physical reservoir computing model based on overdamped bistable stochastic resonance system," Neurocomputing **468**, 137–147 (2022).
[41]A. E. Hramov, N. Kulagin, A. V. Andreev, and A. N. Pisarchik, "Forecasting coherence resonance in a stochastic FitzHugh–Nagumo neuron model using reservoir computing," Chaos, Solitons Fractals **178**, 114354 (2024).
[42]A. Röhm, D. J. Gauthier, and I. Fischer, "Model-free inference of unseen attractors: Reconstructing phase space features from a single noisy trajectory using reservoir computing," Chaos **31**, 103127 (2021).
[43]S. R. Dunbar, *Mathematical Modeling in Economics and Finance: Probability, Stochastic Processes, and Differential Equations* (American Mathematical Society, 2019), Vol. 49.
[44]E. Renshaw, *Stochastic Population Processes: Analysis, Approximations, Simulations* (OUP Oxford, 2015).
[45]E. Platen, "An introduction to numerical methods for stochastic differential equations," Acta Numer. **8**, 197–246 (1999).
[46]P. E. Kloeden and E. Platen, "Stochastic differential equations," in *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin, 2011), pp. 103–160.
[47]P. Billingsley, *Convergence of Probability Measures* (John Wiley & Sons, 2013).
[48]A. Longtin, "Stochastic resonance in neuron models," J. Stat. Phys. **70**, 309–327 (1993).
[49]M. Masoliver, N. Malik, E. Schöll, and A. Zakharova, "Coherence resonance in a network of FitzHugh-Nagumo systems: Interplay of noise, time-delay, and topology," Chaos **27**, 101102 (2017).
[50]E. Baspinar, L. Schülen, S. Olmi, and A. Zakharova, "Coherence resonance in neuronal populations: Mean-field versus network model," Phys. Rev. E **103**, 032308 (2021).
[51]A. N. Pisarchik, A. V. Kir'yanov, Y. O. Barmenkov, and R. Jaimes-Reátegui, "Dynamics of an erbium-doped fiber laser with pump modulation: Theory and experiment," J. Opt. Soc. Am. B **22**, 2107–2114 (2005).
[52]J. Barba-Franco, L. Romo-Muñoz, R. Jaimes-Reátegui, J. García-López, G. Huerta-Cuéllar, and A. N. Pisarchik, "Electronic equivalent of a pump-modulated erbium-doped fiber laser," Integration **89**, 106–113 (2023).
[53]A. S. Pikovsky and J. Kurths, "Coherence resonance in a noise-driven excitable system," Phys. Rev. Lett. **78**, 775 (1997).
[54]N. Platt, E. Spiegel, and C. Tresser, "On-off intermittency: A mechanism for bursting," Phys. Rev. Lett. **70**, 279 (1993).
[55]J. Heagy, N. Platt, and S. Hammel, "Characterization of on-off intermittency," Phys. Rev. E **49**, 1140 (1994).
[56]K. Pyragas, "On-off intermittency at the threshold of weak synchronization of chaos," Chaos, Solitons Fractals **9**, 337–342 (1998).
[57]A. E. Hramov and A. A. Koronovskii, "Intermittent generalized synchronization in unidirectionally coupled chaotic oscillators," Europhys. Lett. **70**, 169 (2005).
[58]O. Moskalenko, A. Koronovskii, and S. Shurygina, "Intermittent behavior at the boundary of noise-induced synchronization," Tech. Phys. **56**, 1369–1372 (2011).
[59]H. D. Abarbanel, N. F. Rulkov, and M. M. Sushchik, "Generalized synchronization of chaos: The auxiliary system approach," Phys. Rev. E **53**, 4528 (1996).
[60]J. A. Platt, A. Wong, R. Clark, S. G. Penny, and H. D. Abarbanel, "Robust forecasting using predictive generalized synchronization in reservoir computing," Chaos **31**, 123118 (2021).

[61] A. E. Hramov, A. A. Koronovskii, and O. I. Moskalenko, Are generalized synchronization and noise-induced synchronization identical types of synchronous behavior of chaotic oscillators?" Phys. Lett. A **354**, 423–427 (2006).
[62] D. Ibáñez-Soria, J. García-Ojalvo, A. Soria-Frisch, and G. Ruffini, "Detection of generalized synchronization using echo state networks," Chaos **28**, 033118 (2018).
[63] N. Frolov, V. Maksimenko, A. Lüttjohann, A. Koronovskii, and A. Hramov, "Feed-forward artificial neural network provides data-driven inference of functional connectivity," Chaos **29**, 091101 (2019).
[64] Z. Lu, B. R. Hunt, and E. Ott, "Attractor reconstruction by machine learning," Chaos **28**, 061104 (2018).
[65] M. Lebedev, "Brain-machine interfaces: An overview," Transl. Neurosci. **5**, 99–110 (2014).
[66] J. A. Roberts, K. J. Friston, and M. Breakspear, "Clinical applications of stochastic dynamic models of the brain, part I: A primer," Biol. Psychiatry: Cogn. Neurosci. Neuroimaging **2**, 216–224 (2017).
[67] J. A. Roberts, K. J. Friston, and M. Breakspear, "Clinical applications of stochastic dynamic models of the brain, part II: A review," Biol. Psychiatry: Cogn. Neurosci. Neuroimaging **2**, 225–234 (2017).
[68] A. E. Hramov, V. A. Maksimenko, and A. N. Pisarchik, "Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states," Phys. Rep. **918**, 1–133 (2021).