

Using machine learning to predict statistical properties of non-stationary dynamical processes: System climate, regime transitions, and the effect of stochasticity

Cite as: Chaos **31**, 033149 (2021); <https://doi.org/10.1063/5.0042598>

Submitted: 31 December 2020 • Accepted: 05 March 2021 • Published Online: 24 March 2021

 Dhruvit Patel, Daniel Canaday, Michelle Girvan, et al.



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **27**, 121102 (2017); <https://doi.org/10.1063/1.5010300>

[Attractor reconstruction by machine learning](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 061104 (2018); <https://doi.org/10.1063/1.5039508>

[Predicting critical transitions in multiscale dynamical systems using reservoir computing](#)

Chaos: An Interdisciplinary Journal of Nonlinear Science **30**, 123126 (2020); <https://doi.org/10.1063/5.0023764>



Chaos

Special Topic: Nonlinear Model Reduction From Equations and Data

Submit Today!

Using machine learning to predict statistical properties of non-stationary dynamical processes: System climate, regime transitions, and the effect of stochasticity

Cite as: Chaos 31, 033149 (2021); doi: 10.1063/5.0042598

Submitted: 31 December 2020 · Accepted: 5 March 2021 ·

Published Online: 24 March 2021



View Online



Export Citation



CrossMark

Dhruvit Patel,^{1,a)} Daniel Canaday,² Michelle Girvan,¹ Andrew Pomerance,² and Edward Ott¹

AFFILIATIONS

¹The Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, Maryland 26742, USA

²Potomac Research LLC, Alexandria, Virginia 22311-1311, USA

^{a)}**Author to whom correspondence should be addressed:** dpp94@umd.edu

ABSTRACT

We develop and test machine learning techniques for successfully using past state time series data and knowledge of a time-dependent system parameter to predict the evolution of the “climate” associated with the long-term behavior of a *non-stationary* dynamical system, where the non-stationary dynamical system is itself unknown. By the term *climate*, we mean the statistical properties of orbits rather than their precise trajectories in time. By the term *non-stationary*, we refer to systems that are, themselves, varying with time. We show that our methods perform well on test systems predicting both continuous gradual climate evolution as well as relatively sudden climate changes (which we refer to as “regime transitions”). We consider not only noiseless (i.e., deterministic) non-stationary dynamical systems, but also climate prediction for non-stationary dynamical systems subject to stochastic forcing (i.e., dynamical noise), and we develop a method for handling this latter case. The main conclusion of this paper is that machine learning has great promise as a new and highly effective approach to accomplishing data driven prediction of non-stationary systems.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0042598>

Prediction of the behavior of dynamical systems that are themselves changing with time or are influenced by independently time evolving external environmental conditions is an important task that appears in many scientific, technological, medical, sociological, and economic settings. Often, knowledge of the scientific principles, underlying equations, and conditions enabling a basis for the construction of models capable of predicting the behavior of such non-stationary systems is lacking or very incomplete. However, one often has available past measurements of the system state as a function of time. Can these kind of data be used to enable prediction of future dynamical system behavior? In this paper, we address this question via an approach utilizing machine learning and find that it can be extremely effective for this purpose.

I. INTRODUCTION/BACKGROUND

Predicting the time evolution of a dynamical system is a central requirement for enabling many applications. Moreover, it is often the case that first principles knowledge of the dynamical system to be predicted is unavailable. Alternatively, one may have access to measured time series data of the previous states of the system. For situations where the system to be predicted is *stationary* in time, machine learning has proven to be effective in using available time series data to accomplish prediction tasks.^{1–25} In this stationary context, particularly for systems with chaotic behavior, it is useful to partition the problem of prediction into (i) short-term prediction of the state of the system and (ii) long-term prediction/replication of the statistical/ergodic dynamical behavior of the system. That

is, even after short-term prediction fails (e.g., due to the sensitive dependence of chaos), we may still desire to replicate the statistical properties of the orbits of the real system. We have in mind the situation in which, following an initial transient, the orbit from an initial condition goes to an “attractor,” i.e., a subset of the state space on which the orbit continues to reside, while, perhaps, evolving chaotically. In such a stationary situation, a measurement of the state at a random instant of time can be thought of as corresponding to a sample taken from an underlying probability distribution, which is invariant under the flow of time. For such systems, the ergodic theorem allows us to characterize the long-term statistical properties of the orbits by equating the time average of a state function over a single typical orbit to the state space average of that function taken with respect to an invariant measure.²⁶ If we succeed in predicting trajectories that have the same long-term statistical properties as those produced by trajectories of the actual system, then we say we have succeeded in replicating the “climate” of that system.

In contrast to the situation described in the above paragraph, we are interested in prediction of *non-stationary* dynamical systems, i.e., systems that themselves vary with time. However, in the case of a non-stationary system, ergodicity no longer holds. Thus, for non-stationary systems, we cannot meaningfully define a representative statistical quantity in the above way,²⁷ and we wish to extend our concept of “climate” to the non-stationary case. To do this, we consider an ensemble of trajectories originating from many randomly chosen initial conditions in the past. At any subsequent given time, after a sufficiently large convergence time has passed since initialization, the ensemble trace out a so-called “snapshot attractor.”^{27–29} This attractor possesses a probability distribution that (under appropriate general conditions) is independent of the set of initial conditions chosen for the ensemble and thus provides an appropriate measure representing the probability distribution of states on the snapshot attractor.²⁷ [The collection of all snapshot attractors (at every time instance) is referred to, in mathematical and terrestrial climate-related literature, as the pullback attractor of the system.³⁰] This time-dependent snapshot probability distribution can then be used to obtain expectation values of state-dependent functions, thereby defining the time-dependent climate of a non-stationary system.

The goal of this paper will be to devise and test machine learning approaches for the prediction of the future evolution of the climate of non-stationary dynamical systems.

We note that, although machine learning has previously been extensively investigated and applied for the task of short-term state prediction of stationary dynamical systems (especially chaotic systems) (see Refs. 1–3 and 13–22) and has also been shown to be effective for replicating the climate of stationary systems,^{1,5,23,24} machine learning has previously received little attention as a general approach to the important task of predicting climate evolution in non-stationary dynamical systems. For recent papers on using machine learning for short-term state prediction of non-stationary systems, see Refs. 19, 25, and 31. In the context of our stated goal, an important issue is that of “regime transitions.” Specifically, for different system parameters, stationary dynamical systems can generally exhibit a variety of different types of motions with

different climates, and, correspondingly, slowly time varying non-stationary systems can experience *regime transitions* where, once the system crosses some critical condition, the system’s behavior rapidly becomes qualitatively and quantitatively different. There are a variety of mechanisms for regime transitions ranging from basic bifurcations of periodic orbits (e.g., Hopf, saddle-node, and period doubling bifurcations) to transitions of chaotic attractors (e.g., period doubling cascades,^{32,33} intermittency,³⁴ and crises^{35,36}). For a very recent work on using machine learning to predict system collapse and chaotic transients associated with crisis transitions, see Ref. 37.

Thus, in addition to slow, temporally continuous climate change, regime transitions are important in many fields. For example, in the context of terrestrial climate, reference is typically made to “tipping points” representing abrupt transitions in the Earth’s climate.^{38,39} Geological records indicate that the Earth has experienced such transitions in the past; for example, Greenland ice-core records indicate a series of abrupt climatic events in the Last Glacial period,⁴⁰ and there is growing concern that anthropogenic activities such as greenhouse gas emissions could drive the climate system to a regime transition,^{41,42} such as the possible shutdown of the thermohaline circulation in the Atlantic Ocean^{43,44} and the potential transition from a wet to a dry Indian summer monsoon.⁴⁵ Such regional transitions could also affect the climate of geographically distant locations via teleconnection.^{46–48} Naturally, both temporally continuous behavior evolution as well as regime change events are of immense consequence not only for atmospheric climate, but also in many other dynamical contexts, e.g., economics,^{49,50} ecology,^{51–62} medicine,^{63,64} technology,^{65,66} and sociology.⁵⁷

The goal of predicting the climate of non-stationary systems from time series state data, in principle, requires some knowledge of the future evolution of the underlying dynamical system that produces the data. Such knowledge might be directly available or derivable from the available past state time series data. The need for this type of knowledge can be seen by considering the example where the underlying system time dependence experiences an abrupt change such that the system after the change is unrelated to the system before the change, thus making predictions of the state changes past the time of the abrupt system change untenable. With this in mind, we restrict our considerations to the case where the underlying system evolution is smooth and slow compared to the time scale (τ_s) of the state evolution ($\tau_{ns} \gg \tau_s$, where τ_{ns} is the “non-stationarity time scale” over which the underlying system evolves). We emphasize that we are interested in climate prediction over the long time scale τ_{ns} . In this paper, we consider the scenario where the non-stationarity of the system to be predicted is due to time-dependent system parameters where the time dependence is known, although the system itself is otherwise unknown. Discussion of the importance of this scenario in application is given in Sec. IV A. (For simplicity, in what follows, we will only consider the case where there is one relevant time-dependent system parameter.)

The main conclusion of this paper is that machine learning offers a surprisingly promising avenue for using measurements of past system state time series to accomplish the important general goal of predicting the long-term dynamical behavior (climate) of

non-stationary dynamical systems, including prediction of apparently continuous change and of more sudden regime transitions.

In what follows, we use reservoir computing^{68–70} as our machine learning platform of choice for practical reasons. Reservoir computing is a computationally inexpensive framework since it requires training only its output layer via a simple linear regression. The resulting reduced training time allows for relatively rapid testing of various methodologies and parameter regimes. However, we expect other types of machine learning, e.g., deep learning,⁷¹ to also work well for this task via the methods we develop and test in this paper.

Another issue we address is that of predicting the time evolution of a non-stationary dynamical system evolving under the influence of dynamical noise, which is often unavoidable in real situations. We present a simple method that uses the residual error distribution from the training procedure to inject appropriately determined noise into the reservoir prediction scheme to perturb the reservoir output in a way that mimics the effects of the dynamical noise present in the real system, thus enabling prediction of the noise-influenced climate.

The rest of the paper is organized as follows. In Sec. II, we give a brief overview of reservoir computing and of a widely employed basic scheme for using it for prediction of stationary dynamical systems. In Sec. III, we discuss noiseless and noisy non-stationary versions of the logistic map and their behaviors. In Sec. IV A, using the noiseless non-stationary logistic map, we introduce and test our method. In Sec. IV B, we generalize the method of Sec. IV A to include dynamical noise. Sections V and VI test our methods on non-stationary versions of the Lorenz 63 system⁷² and the Kuramoto–Sivashinsky system,^{73,74} respectively. Section VII presents conclusions. Referring to Secs. III and VI, we note that the example systems tested include a simple discrete-time map (Sec. III), a low-dimensional continuous-time system (three ordinary differential equations; Sec. V), and a partial differential equation potentially exhibiting spatiotemporal chaos (Sec. VI).

II. REVIEW OF RESERVOIR COMPUTING (RC) AND ITS USE FOR PREDICTION OF STATIONARY SYSTEMS

Reservoir computing (e.g., see the review paper⁶⁸) is a machine learning approach that has proven to be effective in tasks involving dynamical time series data. Although there are various ways of implementing RC,^{75–80} in this paper, we consider the widely used implementation that utilizes a network of neuron-like nodes with recurrent connections, the “reservoir,” to process temporal/sequential data.^{69,70} The main idea is to drive the reservoir with an input signal and form a trainable linear combination of the nonlinear response of the individual network nodes to obtain an output sequence that closely approximates some desired sequence.

For illustrative and background purposes, in this section, we describe the application of reservoir computing to predicting the time evolution of an unknown *stationary* system. Given a set of measurements of the state of a system from some time in the past $t = -T$ up to some time labeled $t = 0$, we desire to predict the state of the system for $t > 0$. Even after prediction of the state of the system fails, we may still desire that the predicted trajectory replicates the climate of the actual system in the long-time limit (Refs. 4 and 5).

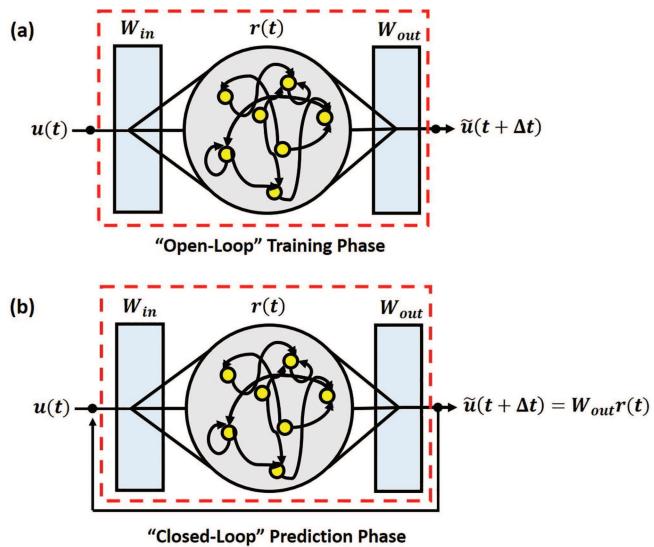


FIG. 1. Schematic of the reservoir computer. (a) The “open-loop” configuration used for training the output weight matrix, W_{out} . An input sequence, $u(t)$, is coupled into a reservoir of nodes via a matrix W_{in} . The excited reservoir state $r(t)$ is mapped to an output, $\tilde{u}(t)$, via a matrix of adjustable weights W_{out} . (b) The “closed-loop” configuration used for prediction. The reservoir output at one time step is fed as input to the reservoir at the next time step. The dashed red box denotes the “reservoir system,” made up of W_{in} , the reservoir, and W_{out} .

We denote the state of the dynamical system (from which the measurements are obtained) at time t by a K dimensional vector $u(t) = [u_1(t), u_2(t), \dots, u_K(t)]^T$. If our reservoir consists of N nodes, then the K dimensional input vector at each iteration is coupled into the reservoir via a $(N \times K)$ dimensional input-to-reservoir coupling matrix, W_{in} (see Fig. 1). We chose to construct this matrix by randomly selecting one element of each row to be a nonzero number chosen randomly from a uniform distribution on the interval $[-\chi, \chi]$. Each reservoir node i has a scalar state $r_i(t)$. The state of the reservoir at time t is denoted by the N dimensional vector $r(t) = [r_1(t), r_2(t), \dots, r_N(t)]^T$. The reservoir evolves dynamically given the linearly coupled input and a network adjacency matrix A , according to

$$r(t + \Delta t) = \tanh(Ar(t) + W_{in}u(t)), \quad (1)$$

where the hyperbolic tangent function is applied to each element of its vector argument. The directed adjacency matrix is a randomly generated sparse matrix such that it has a predetermined average number of connections per node, $\langle d \rangle$. For our task, the desired output time sequence is of the same dimension as the input sequence; therefore, a $(K \times N)$ dimensional reservoir-to-output coupling matrix W_{out} is used to linearly map the reservoir state vector $r(t)$ to the output denoted \tilde{u} at the next time step, $t + \Delta t$. That is, the output is the K -vector, $W_{out}r(t) = \tilde{u}(t + \Delta t)$. These steps are diagrammatically represented in Fig. 1(a). The matrix elements of W_{out} are determined by the training so that the output of the reservoir computer closely approximates the desired output for

$-T \leq t \leq 0$. To do this, we minimize a “cost function,”

$$\sum_{-T \leq t \leq 0} ||W_{out}\mathbf{r}(t) - \mathbf{u}(t + \Delta t)||^2 + \alpha ||W_{out}||^2,$$

where we include a Tikhonov regularization parameter α to prevent overfitting by penalizing large output weights. This minimization is a simple linear regression. The set of coefficients of W_{out} , which minimizes this L^2 norm, is taken to be the “trained” set of output weights for the network. The choice of α and other “hyperparameters” [such as the average degree $\langle d \rangle$ and the spectral radius ρ_s of A (i.e., the largest eigenvalue magnitude of A)] as well as the maximal strength of the input-to-reservoir coupling χ are empirically chosen to minimize the discrepancy between a target prediction and the reservoir prediction over a validation dataset (e.g., by performing a grid search over the space of these hyperparameters). Furthermore, we constrain our search by restricting our attention to a set of hyperparameters that satisfy the “echo state condition” (see, e.g., Ref. 69), which requires that, starting from any two different reservoir states \mathbf{r}_1 and \mathbf{r}_2 , iterations of Eq. (1), after a transient phase, yield the same time series for the reservoir state $\mathbf{r}(t)$; i.e., the reservoir eventually “forgets” its initial condition, meaning that its evolution depends only on its input $\mathbf{u}(t)$ and is independent of its initial state. If N is sufficiently large, it can be expected that the output $\tilde{\mathbf{u}}(t + \Delta t)$ will be extremely close to the desired output $\mathbf{u}(t + \Delta t)$; i.e., the reservoir will give an accurate one time step prediction.

Once we have trained the reservoir on previous measurements of the system states, using the “open-loop” configuration for $-T \leq t \leq 0$ shown in Fig. 1(a), we predict the system evolution for $t > 0$ by using the “closed-loop” configuration shown in Fig. 1(b). In this configuration, the output from the reservoir computer at one step is fed back as input at the next iteration. Thus, on the first iteration, the input is $\mathbf{u}(t)$ and [neglecting the small difference between $\mathbf{u}(t)$ and $\tilde{\mathbf{u}}(t)$] the output is ideally $\mathbf{u}(t + \Delta t)$, which, when fed back, yields an output $\mathbf{u}(t + 2\Delta t)$, which, when fed back, yields an output $\mathbf{u}(t + 3\Delta t)$, and so on. However, due to the small deviation between $\mathbf{u}(t)$ and $\tilde{\mathbf{u}}(t)$, errors build up with multiple circuits of the feedback loop, and eventually, the prediction fails. Nevertheless, it is observed that this scheme often produces predictions that are useful for a limited time. One notable point is that, because the reservoir network is “recurrent” (i.e., has directed links forming closed paths cycling between nodes), the reservoir system retains memory of its past states and inputs.

As mentioned in Sec. I, we are interested in both discrete-time dynamical systems (maps; Secs. III and IV), as well as continuous-time dynamical systems (flows; Secs. V and VI). In the case of maps, t is an integer, where $t = n$ ($n = 0, 1, 2, \dots$), and the reservoir time step is $\Delta t = 1$. In the case of flows, we will choose Δt to be small compared to the characteristic time for variation of $\mathbf{u}(t)$ so that the predicted discrete outputs, separated by Δt , give a good representation of the continuous-time evolution.

It is important to note (e.g., Sec VI) that the closed-loop configuration of Fig. 1(a) is itself a stationary dynamical system whose dynamics, through the training of W_{out} , has essentially been devised so as to mimic the dynamics of the unknown system producing the measurement data (see Ref. 4).

In Sec. IV, we modify this setup (applicable for the prediction of *stationary* dynamical evolution processes) to the application of non-stationary, possibly stochastic, processes.

III. THE NON-STATIONARY LOGISTIC MAP

In this section, we consider the one-dimensional logistic map that provides a simple platform for formulating, illustrating, and testing the application of RC to the task of predicting the time evolution of a non-stationary discrete-time dynamical system. Because of its simplicity, the logistic map allows us to rapidly test various methodologies and techniques. Furthermore, we will subsequently show (Secs. V and VI) that our results and methods, demonstrated for the simple example of the logistic map, carry over to other, less simple systems. Before we demonstrate the reservoir computer’s ability to predict the non-stationary climate of the logistic map, we briefly review three variations of this system: the stationary logistic map, a non-stationary logistic map, and a stochastic non-stationary logistic map. We demonstrate that the dynamics and regime transitions have different characteristics for each of these three variations of this system.

We will first briefly review the behavior of the stationary logistic map system,

$$x_{n+1} = \lambda x_n (1 - x_n), \quad (2)$$

for different but fixed values of λ . We randomly initiate a trajectory in the interval $(0, 1)$, allow it to evolve under the action of the map for fixed λ , and consider only its long-term behavior (after any transients associated with the choice of initial condition have died away). Plotting the iterations of x in this long-term limit for each value of λ , we generate a bifurcation diagram, as shown in Fig. 2(a). For $\lambda \in (1, 3)$, a typical initial condition $x_0 \in (0, 1)$ will evolve to a fixed point attractor, $x^* = 1 - 1/\lambda$. As λ is increased through $\lambda = 3$, the stationary point x^* bifurcates to a period 2 orbit. As λ is increased further, the attracting orbit undergoes a period doubling cascade^{32,33} terminating at $\lambda = \lambda_\infty \approx 3.569 \dots$. In the range $\lambda_\infty < \lambda < 4$, chaos occurs, accompanied by an infinite number of windows of different periodicity; e.g., see Sec. 2.2 in Ref. 81. For this section, in order to conveniently address a variety of regime transition types, we focus on the period 3 window $\lambda \in [3.82, 3.87]$, as shown in Fig. 2(b). As λ increases through a critical value $\lambda_{c_1} \approx 3.828 \dots$, the period 3 window is initiated from a chaotic situation below λ_{c_1} through an intermittency transition³⁴ associated with a period 3 tangent bifurcation. As λ increases beyond λ_{c_2} , this period 3 orbit undergoes a period doubling cascade, leading to chaos cycling through three bands (with interspersed windows). This three-piece chaotic attractor widens for larger λ until it collides with the unstable period 3 orbit generated at the start of the window. This collision results in a crisis transition^{35,82} causing an abrupt broadening of the three-piece chaotic attractor into a single-band chaotic attractor, thereby terminating the period 3 window (at λ_{c_3}).

In the tests and examples in the rest of this paper, we will often consider situations with windows, like the period three window of Fig. 2(b). We emphasize, however, that we do this not necessarily because we want to treat windows, but rather because, as noted above for the case of Fig. 2(b), windows contain within

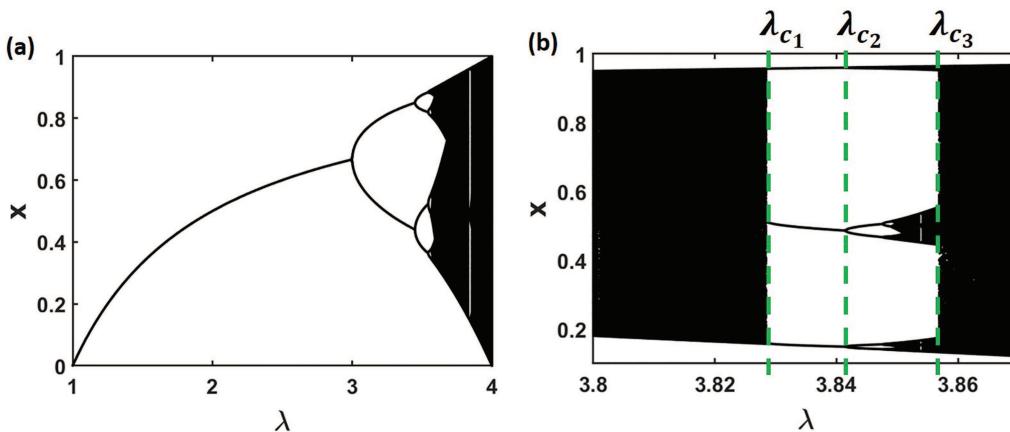


FIG. 2. (a) Bifurcation diagram of the logistic map. (b) Zoomed-in view of the period 3 window of the logistic map bifurcation diagram.

them a variety of general types of common regime transitions (intermittency transitions,³⁴ period doubling cascades, crises^{35,82}). Thus, windows provide a convenient platform for studying and testing the effectiveness of our methods when regime transitions are present.

Now, we modify the stationary system to create a non-stationary situation where the parameter λ drifts linearly with time,

$$x_{n+1} = \lambda_n x_n (1 - x_n), \quad (3a)$$

$$\lambda_{n+1} = \lambda_n + \gamma. \quad (3b)$$

[Later in this paper, we will consider the case where the linear drift with time, Eq. (3b), is replaced by a nonlinear drift with time.] To discuss the dynamics of this system, we consider the situation where λ_n varies very slowly ($|\gamma| \ll \lambda_n$). For $\gamma = 10^{-5}$, Fig. 3(a) shows a typical orbit evolved under the action of Eq. (3) over the interval of the period 3 window. Notice that the period doubling bifurcation, from period 3 to period 6, looks different, more “abrupt” [see blow-up in Fig. 3(b)], in this non-stationary case than in the stationary case shown in Fig. 2(b). This can be understood as follows. Just before the bifurcation, the orbit is approximately moving on the period 3 attractor. When λ_n increases past λ_{c2} , the period 3 orbit points become unstable, but the trajectory remains, for a few iterations, very close to now an unstable period 3 orbit, and then, after a few more time steps, the trajectory becomes sufficiently displaced from the unstable period 3 orbit (since λ_n is changing) and is repelled from the period 3 orbit, quickly moving to the now stable period 6 orbit. Thus, we obtain the observed delayed, but rapid, transition from a period 3 orbit to a period 6 orbit. As the drift rate of λ increases, the transition from a period 3 orbit to a period 6 orbit is further delayed, and the details of the period doubling cascade become less discernable and are eventually washed out, as seen from Figs. 3(c) and 3(d), which apply for $\gamma = 5 \times 10^{-5}$. The main point illustrated in Figs. 3(a)–3(d) is that non-stationarity can qualitatively alter the dynamical behavior of a system, especially during a regime transition.

Similarly, dynamical noise can also significantly alter the climate and features of regime transitions. To illustrate this, we consider the non-stationary stochastic logistic map,⁸³ given by Eq. (4),

$$x_{n+1} = \lambda_n x_n (1 - x_n) + \varepsilon_n, \quad (4a)$$

$$\lambda_{n+1} = \lambda_n + \gamma, \quad (4b)$$

where ε_n (the dynamical noise) is a number randomly and independently chosen from a distribution $\zeta(\varepsilon)$. For simplicity, we will consider the case in which the noise distribution is uniform on the interval $[-\varepsilon, \varepsilon]$. As shown in Ref. 84 for the stationary system bifurcation diagram, the dynamical noise acts to blur the fine structure of the period doubling cascade (near λ_∞) and appears to cause the period doubling cascade to terminate after a few period doublings. Turning to the non-stationary case, in Fig. 3(e), we have plotted a typical trajectory in the period 3 window evolving under the action of Eq. (4) with a noise of strength $\varepsilon = 0.002$ and λ drifting linearly at a rate $\gamma = 10^{-5}$. We see that the dynamical noise “blurs” the period doubling bifurcation from a period 3 orbit to a period 6 orbit. We also see, before the occurrence of the internal crisis, at the upper λ range of the window, that the noisy trajectory intermittently leaves (and returns) to the region of the state space, which corresponds to the three-piece chaotic attractor of the stationary system. Thus, dynamical noise can cause a significant change in features of regime transitions, including blurring of the noiseless sharp location of the point at which a transition occurs.

We desire to accurately predict and replicate, from past time series data, the effects of both noiseless and noisy non-stationarity on the climate associated with time evolution of the system.

IV. APPLICATION OF RC TO CLIMATE PREDICTION OF THE NON-STATIONARY LOGISTIC MAP

A. The noiseless non-stationary logistic map

Now, we turn to our stated goal: given a finite set of measurements of the state of an unknown non-stationary dynamical system

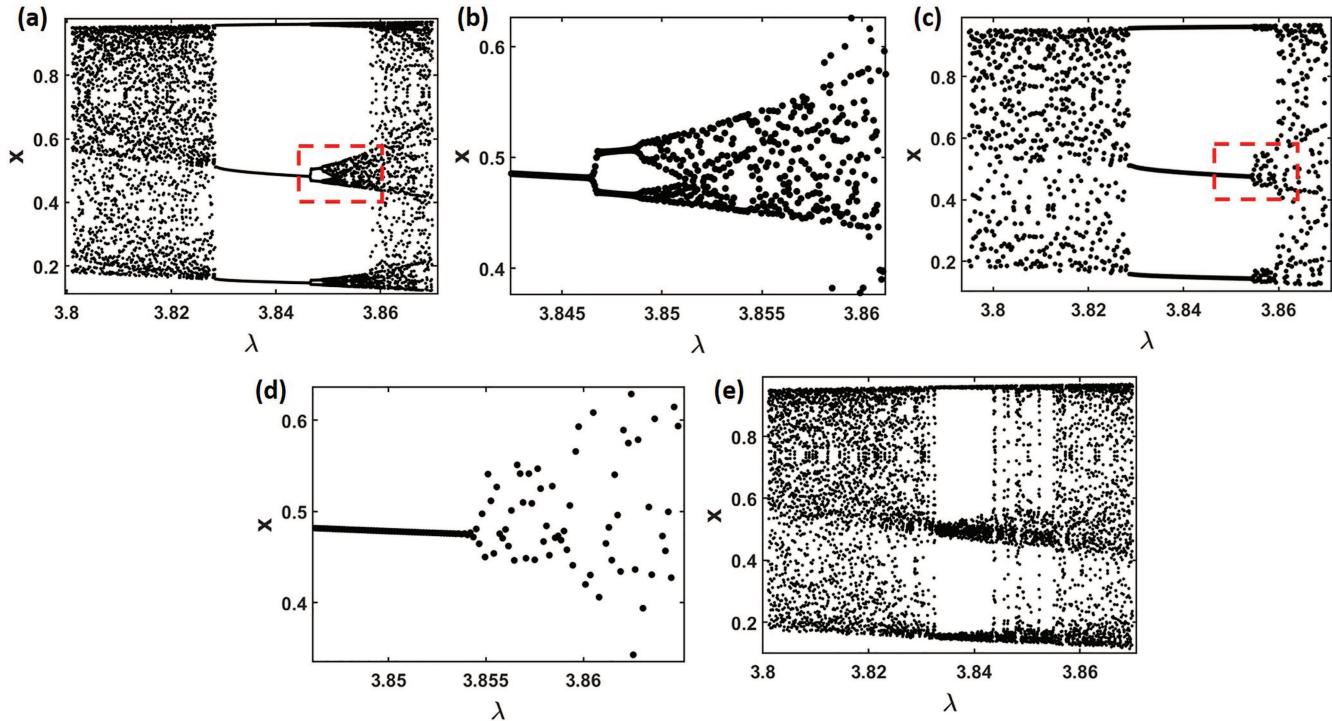


FIG. 3. A trajectory evolving under the non-stationary logistic map (a) given by Eq. (3) for $\gamma = 10^{-5}$, (c) given by Eq. (3) for $\gamma = 5 \times 10^{-5}$, and (e) given by Eq. (4) for $\gamma = 10^{-5}$ and $\varepsilon = 0.002$. (b) and (d) show a zoomed-in view of the red boxed areas in (a) and (c), respectively, near the period doubling bifurcation from a period 3 orbit to a period 6 orbit.

from some time in the past $t = -T$ up to some time $t = 0$, we predict the climate and regime transitions for $t > 0$. More precisely, when the prediction is performed over an ensemble of trajectories, each initiated from a randomly chosen initial condition, we want our predicted ensemble of trajectories to replicate the non-stationary climate of the actual system. For work previously applying initial condition ensembles for studying and understanding non-stationary dynamics, see Refs. 85–87. First, we demonstrate how naively applying RC using the measurements of the previous states of the system as the only input to the reservoir (i.e., following the procedure of Sec. II) fails to accomplish our goal. We will call this setup “Setup 1.” Considering the noiseless system given by Eq. (3) for $\gamma = 10^{-5}$, we train the reservoir in the “open-loop” configuration [Fig. 1(a)] using a set of training data, $\{x_n\}$ for $n \leq 0$ (1000 data points) and then allow the reservoir to autonomously predict a trajectory, x_n for $n > 0$, forward in time using the “closed-loop” configuration [Fig. 1(b)]. Table I displays the set of hyperparameters used in all tasks in this paper involving the prediction of the logistic map. Figure 4(a) compares the actual trajectory of the system (black dots) and the reservoir-predicted trajectory (red dots). We see that the reservoir-predicted trajectory does not undergo any of the regime transitions that the actual trajectory undergoes. Alternatively, the reservoir-predicted trajectory appears to evolve according to a fixed stationary map. Going back to the description of the prediction in the stationary case, Sec. II, this is to be expected because the memory

of the reservoir is short compared to the timescale for change of λ_n (i.e., γ^{-1}), and the training produces a single fixed W_{out} , which, when used in the closed-loop prediction configuration, produces a stationary reservoir dynamical system. As such, we can think of the training procedure of Fig. 1(a) acting on non-stationary training data as choosing a W_{out} that in some sense yields the stationary system that is most consistent with the non-stationary training data $-T < t < 0$. Figure 4(b) shows the map traced out by the reservoir-predicted trajectory (red dots) and by the actual trajectory (black dots) from Fig. 4(a).

In order to take into account non-stationarity, we must break the stationarity of the closed-loop dynamical system of Fig. 1.

TABLE I. Hyperparameters for predicting the logistic map.

Reservoir size	N	500
Average number of connections per node	$\langle d \rangle$	1
Spectral radius of the reservoir’s adjacency matrix	ρ_s	0.55
Maximal coupling strength of the input to the reservoir	χ	0.23
Tikhonov regularization parameter	α	2.15×10^{-11}

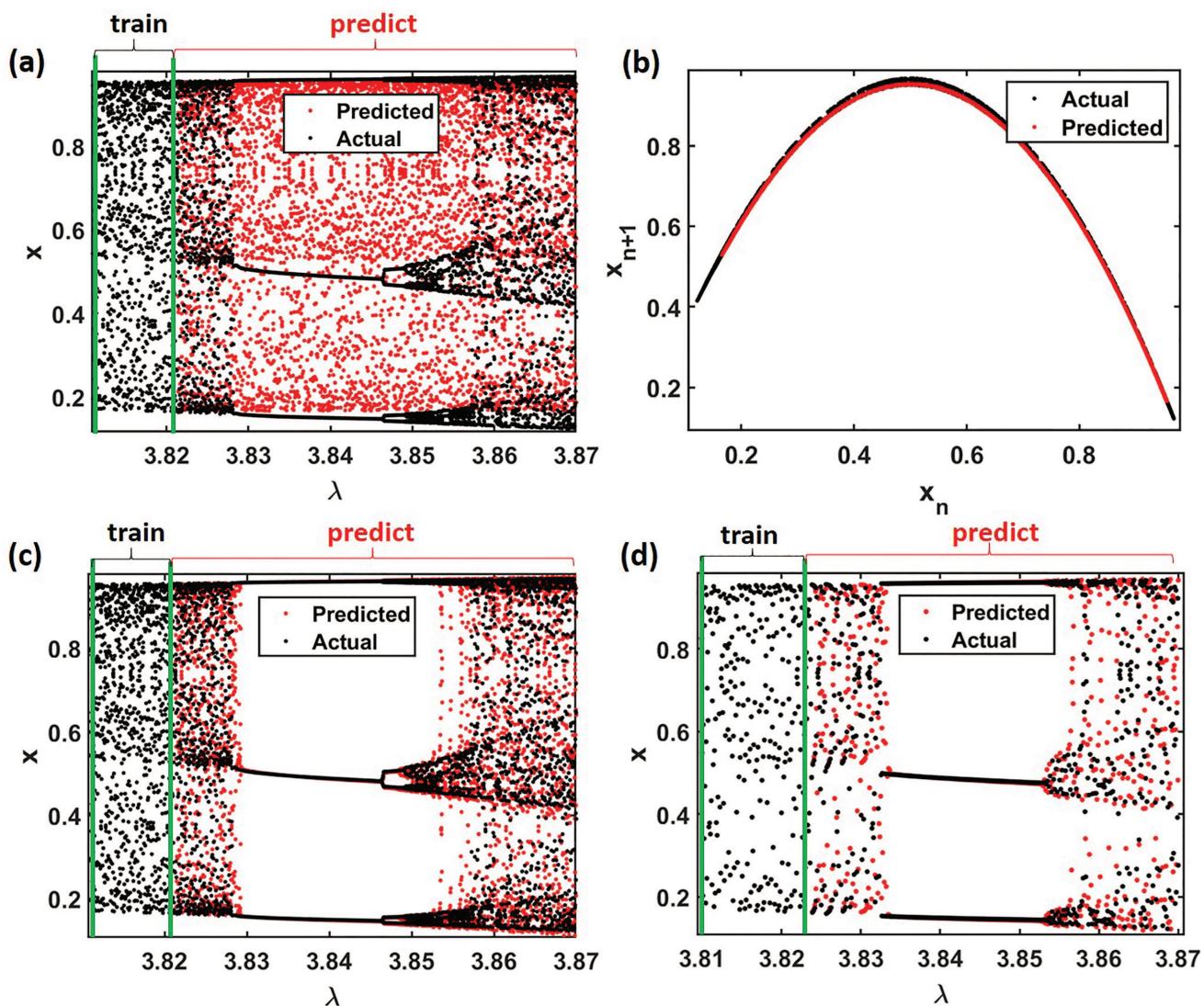


FIG. 4. (a) Comparison of the actual (black dots) and the reservoir-predicted (red dots) trajectories of the logistic map for Setup 1. (b) The logistic map for $\lambda = 3.8$ (black) and the map obtained from the reservoir-predicted trajectory (red dots) of (a). (c) and (d) compare the actual and the reservoir-predicted trajectories for $\gamma = 10^{-5}$ and $\gamma = 5 \times 10^{-5}$, respectively, for Setup 2.

Instead of providing only the state of the dynamical system as input to the reservoir, we also provide the reservoir with the value of the driving parameter λ_n of the system during both the training and the prediction. We call this setup, illustrated in Fig. 5, “Setup 2.” (A similar scheme was very recently independently formulated in the pre-print Ref. 37 for studying crises.) We have coupled this parameter to the reservoir in the same way we couple the measured states of the dynamical system; i.e., each reservoir node is coupled to a single randomly selected element of the state vector $u(t)$ or to the driving parameter. Using Setup 2, we will now demonstrate the successful application of RC to the prediction of the noiseless non-stationary

logistic map. With an additional modification to this method, we will later extend its capabilities to successfully predict the stochastic logistic map as well.

We assume that we have a set of measurements of the state of the system, as well as knowledge of the time variation of the parameter, $\{x_n, \lambda_n\}_{n=-T}^{n=0}$, from some time in the past $n = -T$ to the present $n = 0$, and we desire to predict the future climate of this non-stationary system. We note that the task of prediction where non-stationarity is due to time variation of a *known* parameter is often of great practical interest. For instance, suppose we are trying to predict how the Earth’s climate will change over time, and

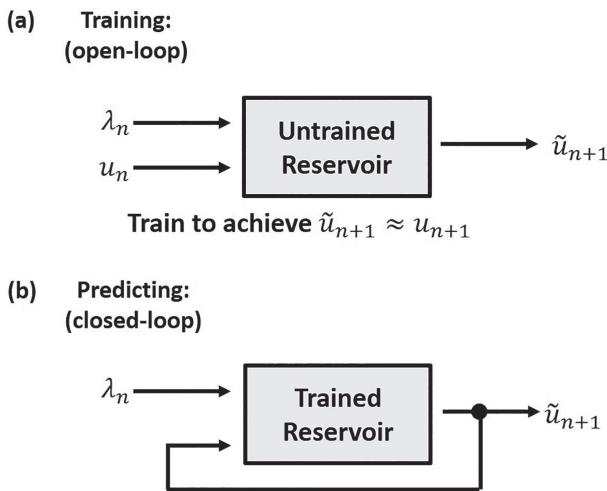


FIG. 5. (a) Open-loop training phase and (b) the closed-loop prediction phase of Setup 2. The rectangular boxes, as for the dashed rectangles in Figs. 1(a) and 1(b), contain the RC system consisting of the input layer (W_{in}), the reservoir, and the output layer (W_{out}).

we have a set of measurements that describe the climate from the past as well as of relevant quantities that we think may directly influence the dynamical behavior of the terrestrial climate system, such as atmospheric CO₂ concentration.^{41,42} We may be interested in forecasting how the climate would change in the future for different rates of increase of CO₂ concentration in the atmosphere. In that case, we may treat CO₂ concentration levels as a known time-dependent driving parameter of the system. With this in mind, we train the reservoir, as shown in Fig. 5(a), and then we employ the “closed-loop” configuration of Fig. 5(b) to predict the future states of the system. For the case with $\gamma = 10^{-5}$, $\varepsilon = 0$, the results are shown in Fig. 4(c). The black dots represent the actual trajectory evolving under Eqs. (3a) and (3b), while the red dots represent the trajectory predicted by the reservoir computer. The reservoir computer was trained on a portion of the actual trajectory corresponding to the interval $\lambda \approx [3.81, 3.82]$ ($T = 1000$ data points). Although the reservoir quickly loses the ability to predict the precise time evolution of the true system orbit, as expected for a chaotic orbit, we see that it was able to replicate and predict the climate, including the various regime transitions: the intermittency transition near $\lambda = 3.828$ from a chaotic orbit to a stable period three orbit, the cascade of period doubling bifurcations starting at $\lambda \approx 3.845$ leading to chaos near $\lambda = 3.857$, and the crisis transition at the upper boundary of the period 3 window ($\lambda \approx 3.856$). Notice that the reservoir-predicted trajectory was able to accurately capture the effects of non-stationarity, for example, in the sudden transition from a period 3 orbit to a period 6 orbit discussed previously. We have also considered the reverse bifurcations (obtained as λ_n decreases with time), and they are well-predicted by the machine learning method; however, we omit these figures here but will show the corresponding figures for the next example (the Lorenz system, Sec. V). Figure 4(d) shows a similar result, but here, the system is

drifting at a faster rate ($\gamma = 5 \times 10^{-5}$). We see that this faster drift alters the details of the transition quite a bit (e.g., we cannot clearly see a period doubling cascade), while, the reservoir, this time trained on $T = 250$ data points (over a similar range of λ as for the previous case), again accurately mimics the dynamics.

We can, more quantifiably, verify that the long-term dynamical behavior predicted by the reservoir computer is indeed characteristic of that of the true system as follows. We initiate a trajectory x_n from an initial condition randomly selected from the interval (0, 1). We train the reservoir over a portion of the actual trajectory and then predict forward in time. We repeat this many times, each time using a trajectory initiated from a different randomly chosen initial condition. [The number of iterations of this step is made large enough that Γ (defined below) appears to be well converged; i.e., Γ does not change appreciably when there is an increase in the number of iterates.] After prediction has started, for each of the actual and predicted trajectories, we collect those points in the trajectory, which are contained in the segment of the trajectory where $\lambda \in [\lambda^*, \lambda^* + \Delta\lambda]$ for a predetermined small $\Delta\lambda$ and a set of λ^* . For each such interval, we then record the orbit points for the actual and predicted trajectories from which we approximate the probability distribution functions of x on the intervals $[\lambda^*, \lambda^* + \Delta\lambda]$. This allows us to compare the x -distributions from the true system and from the RC prediction for different λ^* values. To assist in doing this, we define the distance between the two probability distributions as

$$\Gamma(\lambda) = \frac{2}{\Delta_x} \int |f_a(x; \lambda) - f_p(x; \lambda)| dx, \quad (5)$$

where $f_a(x; \lambda)$ is the approximation to the cumulative probability distribution of x over a small interval $[\lambda^*, \lambda^* + \Delta\lambda]$ obtained from the ensemble of actual trajectories, $f_p(x; \lambda)$ is that which is obtained from the ensemble of reservoir-predicted trajectories, and $\Delta_x = x_{\max} - x_{\min}$ is the range in x of the attracting set. For each portion of the trajectory of the system for which we calculate Γ , the corresponding Δ_x is determined by taking the difference between the minimum and maximum values of the map points, $x_{\max} - x_{\min}$, over the entire ensemble of trajectories, which provides a good estimate of the size of the attractor over that portion of the trajectory. Without the normalization factor of $2/\Delta_x$, this metric is called the Wasserstein metric,⁸⁸ a notion of distance between the two probability distributions. We choose to use this metric because it accounts for the underlying metric space from which the distributions are drawn and also because, since it uses cumulative distributions rather than histograms, it does not depend on an (arbitrary) choice of histogram binning width. This latter property is of particular importance in the analysis we perform. As an example, consider the situation of predicting an orbit in the periodic regime. If the reservoir-predicted orbit accurately captures the periodicity of the true orbit, but the predicted orbit values are slightly displaced from the true orbit values, then depending on the choice of the histogram binning width, a statistical distance that directly compares the predicted and true probability distributions (e.g., the Kullback–Leibler divergence⁸⁹) may suggest that the prediction was either very good or very bad. This arbitrariness is absent when comparing the underlying cumulative probability distributions using the Wasserstein metric. An intuitive interpretation of the Wasserstein metric is as the minimum

“work” needed to transform one probability distribution into the other, where we imagine that the distributions are approximated by histograms with very narrow bins, each bin containing very many samples, and that the transformation from one distribution to the other is accomplished by shifting samples to different bins, in terms of which, the work is then defined as the average over all samples of their x -length shifts. We added the normalization factor $\Delta_x/2$ in our definition of Γ because in general, the integral has units that are the same as those of x and because we wish a small value of the metric Γ to have meaning independent of Δ_x [e.g., if we are considering the logistic map vs the Lorenz system (Sec. V)]. Figure 6(a) shows $\Gamma(\lambda)$, calculated using an ensemble of 2000 randomly initiated trajectories, at locations throughout the period 3 window for the case with $\gamma = 10^{-5}$. We see that there are two instances where $\Gamma(\lambda)$ peaks sharply. The first spike, near $\lambda \approx 3.829$, occurs when the actual system undergoes a transition from chaos to period 3 and the reservoir fails to predict the location of this transition. Note, however, that the narrow width of this spike indicates that this error of the transition point is fairly small. The second spike occurs when the reservoir fails to predict exactly when the internal crisis leads to the abrupt broadening of the three-piece chaotic attracting set into a single-band chaotic set near $\lambda \approx 3.860$. Figure 6(b) shows plots of cumulative probability distributions of the true system (solid curves) and for the RC prediction (dashed curves) at several values of λ indicated in Fig. 6(a) (corresponding to $\lambda_1^* = 3.825$, $\lambda_2^* = 3.845$, $\lambda_3^* = 3.865$, and $\Delta\lambda = 0.001$). For those cases, we calculated $\Gamma(\lambda_1^*) = 0.0211$, $\Gamma(\lambda_2^*) = 0.0064$, and $\Gamma(\lambda_3^*) = 0.0245$. Excellent climate predictions are indicated by the close agreement of the solid and dashed curves in Fig. 6(b) and the small values of Γ in Fig. 6(a).

We now consider the case where the time evolution of λ_n is not linear,

$$x_{n+1} = \lambda_n x_n (1 - x_n), \quad (6a)$$

$$\lambda_{n+1} = \lambda_0 + \lambda_1 \sin(\theta_n) + m(\theta_n/\gamma), \quad (6b)$$

$$\theta_{n+1} = \theta_n + \gamma. \quad (6c)$$

Figure 7(a) shows how the slowly changing parameter λ_n varies in time for $\theta_0 = -\pi/8$, $\gamma = 0.001$, $\lambda_0 = 3.79$, $\lambda_1 = 0.03$, and $m = 10^{-5}$ [note that for this case, $0 < \lambda_1 < (m/\gamma)$, and the variation of λ_n with time n is non-monotonic]. Figure 7(b) compares a trajectory evolving under the action of this map to one predicted by the reservoir computer, again using the method illustrated in Fig. 5. Here, we have trained the reservoir computer on a set of data ($T = 1000$) where λ is changing almost linearly [see Figs. 7(a) and 7(b)]. After this training segment, we see that λ_n deviates more strongly from linear behavior, and remarkably, the reservoir computer can predict the corresponding changes to the nature of the dynamics of the trajectory. We see that the reservoir computer is able to predict regime transitions and dynamics which it was not exposed to during its training. After training, initially, λ continues to increase through λ_{c1} and the dynamics experience a transition from chaotic to periodic via a tangent bifurcation, but shortly after, λ begins to decrease and the system undergoes an intermittency transition in the reverse direction (i.e., from periodic to chaotic) as λ decreases below λ_{c1} . After this reverse transition, λ rapidly

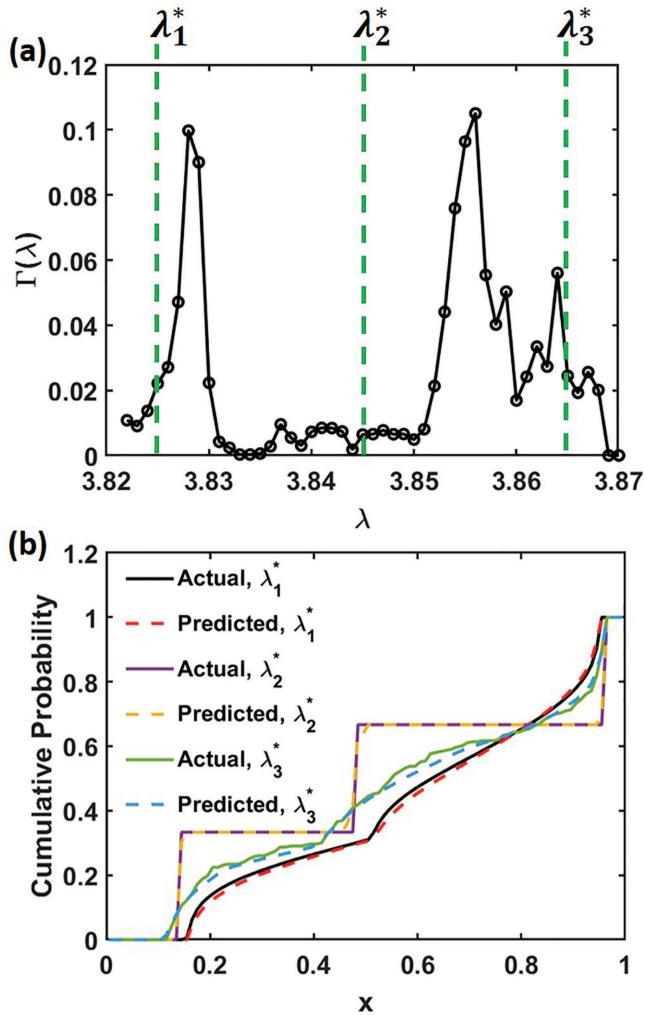


FIG. 6. (a) $\Gamma(\lambda)$ for a set of λ in the period 3 window and $\Delta\lambda = 0.001$. (b) Comparison of the approximate cumulative probability distributions obtained from real trajectories (solid black, purple, and green curves) and predicted trajectories (dashed red, yellow, and blue curves) for $\lambda_1^* = 3.825$, $\lambda_2^* = 3.845$, and $\lambda_3^* = 3.865$ [vertical dashed lines in (a)]. The numerical approximations to $f_a(x, \lambda)$ and $f_p(x, \lambda)$ are taken as the fraction of sample values less than x . The agreement between the actual and predicted cumulative distributions indicates that the reservoir-predicted trajectories accurately represent the dynamics of the different regimes (i.e., the pre-transition chaos, the periodic orbit, and the post-transition chaos).

increases and traverses the range of the period 3 window, which results in an intermittency transition from chaotic below λ_{c1} to periodic above it, and then an abbreviated period doubling cascade starting near λ_{c2} and a crisis transition from a three-piece chaotic attractor to a single-band chaotic attractor near λ_{c3} . The reservoir computer successfully predicts the multiple forward and backward regime transitions experienced by the actual system in this window, as shown in Figs. 7(b) and 7(c). Figure 7(d) shows $\Gamma(\lambda)$, obtained by using an ensemble of 2000 randomly initiated

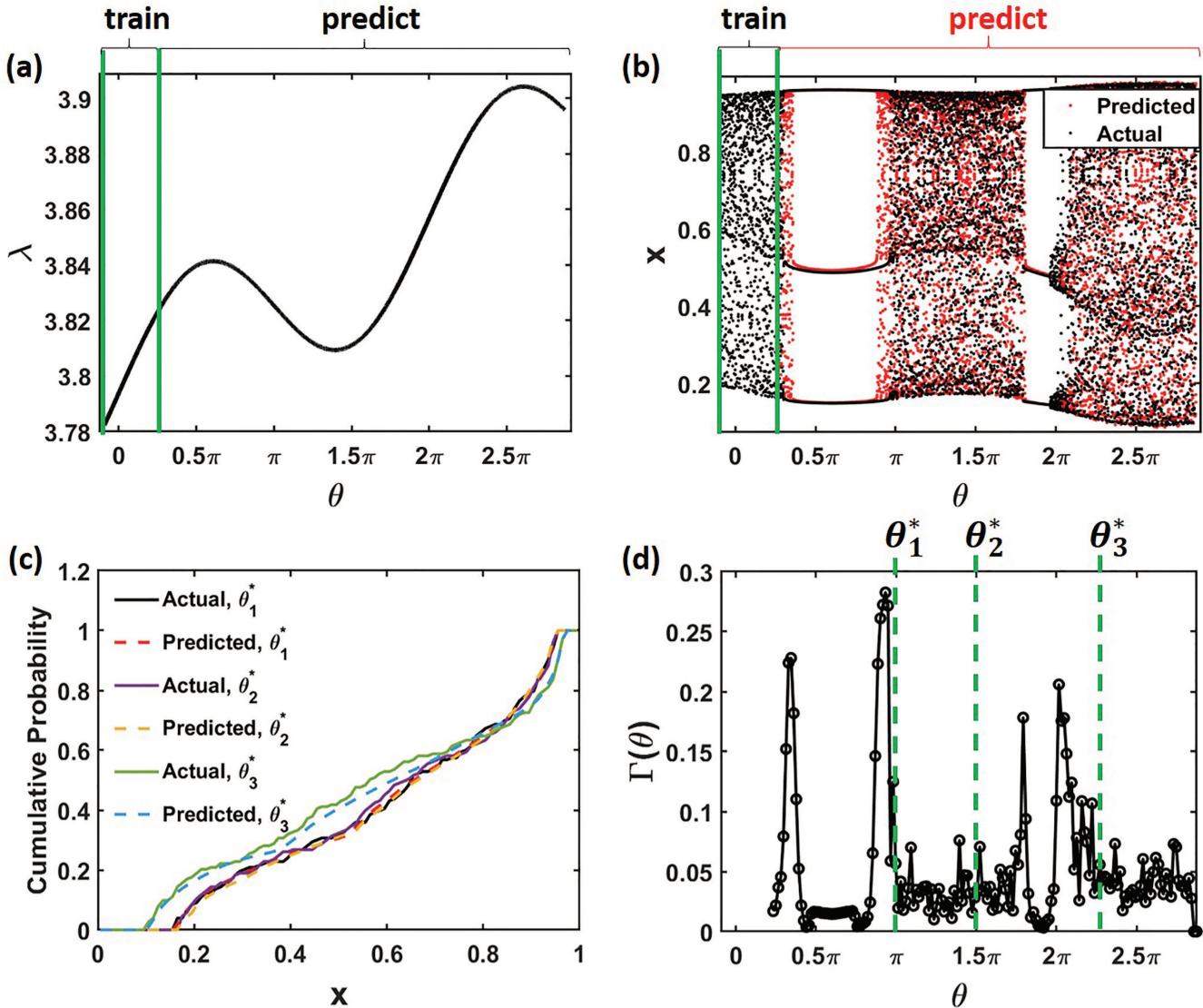


FIG. 7. (a) Time dependence of λ . (b) Comparison of an actual trajectory of the system (black dots) and the corresponding predicted trajectory (red dots). (c) Comparison of the approximate cumulative probability distributions obtained from real trajectories (solid black, purple, and green curves) and predicted trajectories (dashed red, yellow, and blue curves) for $\theta_1^* = \pi$, $\theta_2^* = 1.5\pi$, and $\theta_3^* = 2.25\pi$ ($\lambda_1^* = 3.8253$, $\lambda_2^* = 3.8111$, and $\lambda_3^* = 3.8858$, respectively) corresponding to the vertical green dashed lines in (d). (d) $\Gamma(\theta)$ over the prediction interval.

trajectories and the corresponding reservoir predictions, over the entire prediction interval. Once again, we see local maxima in $\Gamma(\lambda)$ at points near a regime transition, and, as before, this can be attributed to the reservoir predicting a transition slightly earlier or later than it actually occurs. Figure 7(c) explicitly compares $f_a(x; \lambda)$ and $f_p(x; \lambda)$ for $\theta_1^* = \pi$, $\theta_2^* = 1.5\pi$, $\theta_3^* = 2.25\pi$, and $\Delta\theta = 0.05$ ($\lambda_1^* = 3.8253$, $\lambda_2^* = 3.8111$, and $\lambda_3^* = 3.8858$, respectively). The calculated errors between the distributions are small $\Gamma(\theta_1^*) = 0.0195$, $\Gamma(\theta_2^*) = 0.0342$, and $\Gamma(\theta_3^*) = 0.0390$, indicating good prediction of the climate at these points.

B. The noisy non-stationary logistic map

Next, we consider the noisy (or stochastic) logistic map, given by Eq. (4). Once again, we assume that we have a set of measurements of x_n and knowledge of the time variation of the parameter, $\{x_n, \lambda_n\}_{n=-T}^{n=0}$ from some point in the past ($n = -T$) up to the present ($n = 0$), and we wish to predict the future trajectory of the system ($n > 0$). Applying the method of Sec. IV A (Setup 2, Fig. 5) leads us to predicted trajectories, which do not accurately capture the noisy dynamics of the true system. Figure 8 shows this result for two different noise levels (λ is taken to drift linearly with $\gamma = 10^{-5}$).

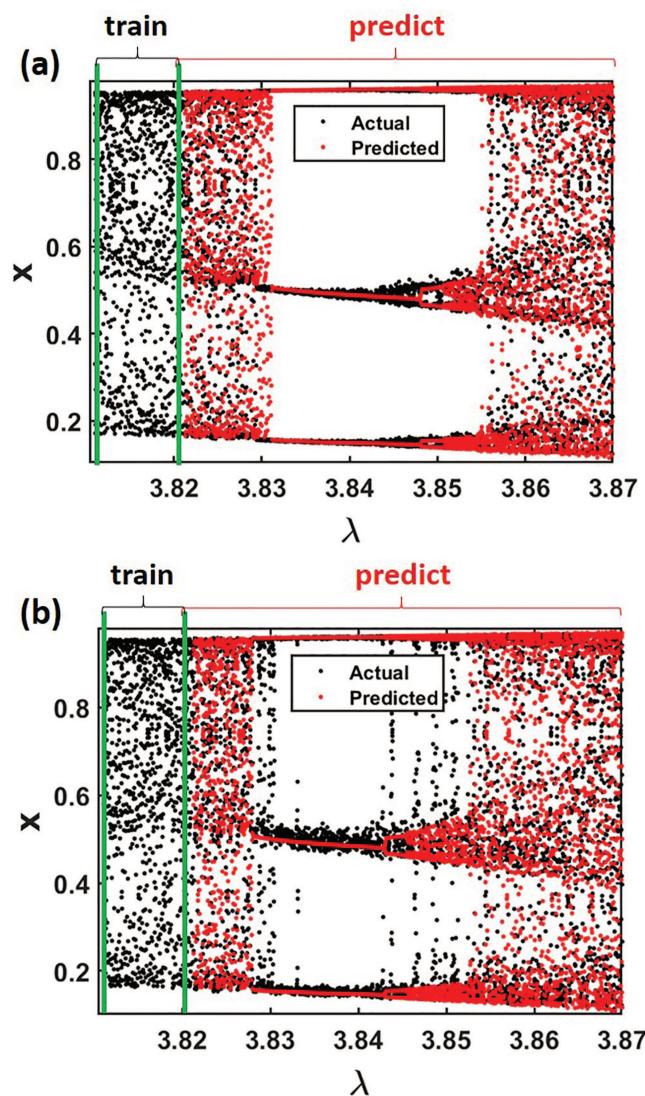


FIG. 8. A comparison of trajectories evolving under the noisy logistic map (black dots) and those predicted by the reservoir (red dots that overlay the black dots) for two noise levels (a) $\varepsilon = 0.0005$ and (b) $\varepsilon = 0.002$.

Note, in particular, the lack of spread in the red dots (as opposed to the black dots) in the period 3 range of λ , as well as the actual system's (black dots) intermittent bursting away from the three chaotic attractor pieces preceding the interior crisis at the upper- λ boundary of the window (which, as seen from the red dots, for the prediction, occurs abruptly, without being preceded by bursting).

We see that despite training the reservoir computer using a dataset obtained from a noisy trajectory, in agreement with Ref. 90, we obtain a prediction that appears to approximate a *noiseless* trajectory. This is because the linear regression used in the training

(see Sec. II) fits the W_{out} elements to many data pairs (x_n, x_{n+1}) , thus averaging out the dynamical noise component (which has zero mean). That is, at time step n , if the input to the trained reservoir is x_n , then its output is an approximation to $\lambda_n x_n (1 - x_n)$, the result of applying the *noiseless* map to x_n . To restore the effect of dynamical noise on the climate, we begin by representing the true noisy orbit by introducing an intermediate variable v_n ,

$$v_{n+1} = \lambda_n x_n (1 - x_n) \quad (7)$$

in terms of which, by Eq. (4a), the true noisy value of x_n is $x_{n+1} = v_{n+1} + \varepsilon_n$, where ε_n (the noise) is a random number with probability distribution $\zeta(\varepsilon_n)$. This can be diagrammatically represented as shown in Fig. 9(a), where an input x_n is mapped according to the noiseless logistic map to v_{n+1} . Based on this, we modify our previous non-stationary reservoir prediction scheme [Fig. 5(b)] by adding a stochastic noise component, as illustrated in Fig. 9(b), where \tilde{v}_{n+1} is the reservoir approximation to v_{n+1} and $\tilde{\varepsilon}_n$ is a random independent variable with probability distribution $\zeta(\tilde{\varepsilon}_n)$, where $\tilde{\zeta}$ is our approximation to the actual noise probability distribution ζ . Given the training data time series $\{x_n\}$ used to train the RC system, we obtain $\tilde{\zeta}$, our approximation to ζ , as follows. Inputting λ_n and x_n to the trained RC system, we obtain the corresponding \tilde{v}_{n+1} for each n in the training time interval. Next, we determine $\tilde{\varepsilon}_n$ via $\tilde{\varepsilon}_n = x_{n+1} - \tilde{v}_{n+1}$, where x_n comes from the training data. We then make a histogram of this collection $\{\tilde{\varepsilon}_n\}$ of values and take this histogram to be our approximate noise probability distribution $\tilde{\zeta}$.

Once we have determined a suitable $\tilde{\zeta}(\tilde{\varepsilon})$, we use it to generate samples $\tilde{\varepsilon}_n$ for n in the prediction time range for use in the prediction phase, Fig. 9(b). Figure 10 shows results for two different noise levels with a linearly drifting λ_n ($\gamma = 10^{-5}$), again demonstrating good climate prediction.

Since $\varepsilon_n = x_{n+1} - v_{n+1}$ and $\tilde{\varepsilon}_n = x_{n+1} - \tilde{v}_{n+1}$, the accuracy of our numerical approximation of the distribution ζ by $\tilde{\zeta}$ depends on the accuracy of the reservoir's approximation \tilde{v}_{n+1} to the actual, noiseless system response v_{n+1} . In the Appendix, we discuss this issue further.

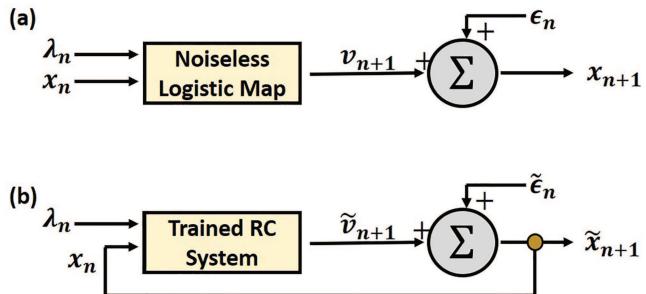


FIG. 9. (a) Diagrammatic representation of the action of the noisy logistic map, using the intermediate variable v_n . (b) Diagram depicting how to approximate the dynamics of the noisy logistic map using a trained reservoir.

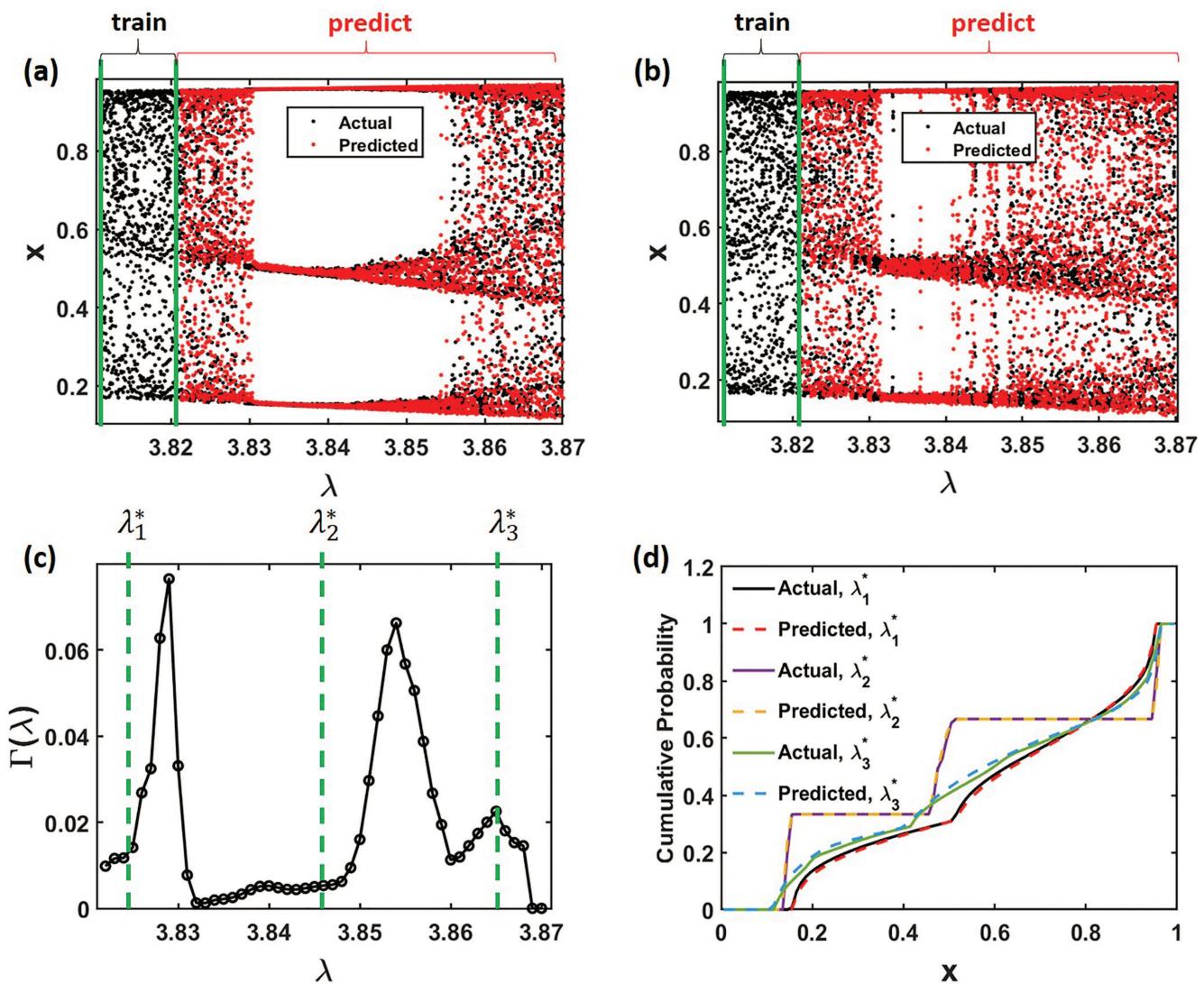


FIG. 10. (a) and (b) Comparison of trajectories evolving under the noisy logistic map (black dots) and those predicted by the reservoir (red dots that overlay the black dots) operating under the modified scheme with noise-injection [Fig. 9(b)] for two noise levels in the actual system (a) $\varepsilon = 0.0005$ and (b) $\varepsilon = 0.002$. (c) $\Gamma(\lambda)$ over the prediction interval. (d) Approximate cumulative probability distributions from actual trajectories (solid curves) and predicted trajectories (dashed curves) for $\lambda_1^* = 3.8325$, $\lambda_2^* = 3.845$, and $\lambda_3^* = 3.865$ and $\Delta\lambda = 0.001$ for $\varepsilon = 0.002$.

V. APPLICATION OF RC TO PREDICTING THE LORENZ 63 SYSTEM

Having shown the reservoir computer's ability to predict climate evolution and regime transitions in a simple discrete-time one-dimensional map, we now implement this approach for the continuous-time three-dimensional Lorenz system,⁷²

$$\frac{dx}{dt} = \sigma(y - x), \quad (8a)$$

$$\frac{dy}{dt} = x(\rho - z) - y, \quad (8b)$$

$$\frac{dz}{dt} = xy - \beta z. \quad (8c)$$

For certain choices of the parameters σ , ρ , and β , typical initial conditions on the attractor of this system give rise to chaotic trajectories. Previous work shows the ability of reservoir computing to predict the short-term trajectories, as well as the long-term climate, of this system when the parameters are fixed.^{1,3–5,18} Here, we expand

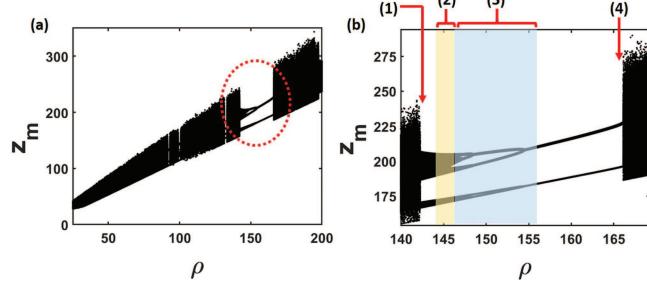


FIG. 11. (a) Lorenz system bifurcation diagram for the interval $\rho = [25, 200]$. (b) A zoomed-in view of the window in the dotted red oval of (a): (1) indicates the crisis transition from a one-piece chaotic attractor to a two-piece chaotic attractor, (2) (and the translucent yellow region) indicates the region of pairwise chaotic band splittings, (3) (and the translucent blue region) shows the region of a period halving cascade, and (4) indicates the intermittency transition from a period 2 orbit to chaos.

on those findings to show that we can predict the long-term dynamical behavior of this system when its parameters are drifting, making the trajectories statistically nonstationary.

We consider time varying ρ in Eq. (8b) while fixing $\sigma = 10$ and $\beta = 8/3$. To display how the system climate evolves as ρ varies, we take a Poincaré surface of section specified by $dz/dt = xy - \beta z = 0$ and $d^2z/dt^2 < 0$ (i.e., we consider maxima of z). We first construct a stationary system bifurcation diagram as follows. For each $\rho_n = n\Delta\rho$ and $\Delta\rho = 0.01$ in the interval $[25, 200]$, we allow a trajectory to evolve under this system and observe its stationary behavior (i.e., with ρ_n constant in time). Figure 11(a) shows the resulting plot of the z -coordinate maxima, z_m , vs ρ , where $z_m \equiv z(t_m)$ and t_m denotes the time of the m th piercing of the surface of section $[x(t_m)y(t_m) = \beta z(t_m)]$ for $d^2z/dt^2 < 0$ at $t = t_m$. In applying RC to predict the time evolution of this system, we restrict our attention to ρ in the range $[140, 170]$. In this range [see Fig. 11(b)], there are eight types of regime transitions, four as ρ increases and four corresponding reverse transitions as ρ decreases. The transitions as ρ increases are (1) the crisis near $\rho \approx 142$ where the one-piece chaotic attractor splits into a two-piece chaotic attractor, (2) a cascade of pairwise chaotic band splittings beginning near $\rho \approx 145$, (3) a period halving cascade beginning near $\rho \approx 146.5$, and (4) an intermittency transition from a period 2 orbit to chaos near $\rho \approx 166$.

We first demonstrate the reservoir's ability to predict the four transitions as ρ increases linearly according to $\rho(t) = \rho_0 + \gamma t$, where $\gamma > 0$ and ρ_0 is an initial value of ρ . We chose $\gamma = 0.1$. As before with the logistic map (Fig. 5), we assume that we have available to us a set of measurements of the system's state as well as knowledge of the time variation of the parameter ρ , $\{u(t), \rho(t)\}_{t=-T}^{t=0}$ from some point in the past $t = -T$ (taken as $T \cong 170$ for this example) up to the “present” $t = 0$ in increments of Δt dictated by the sampling rate of the measurement process. We will use $\Delta t = 0.01$ in the examples that follow. The state of the system is represented by $u(t) = [x(t), y(t), z(t)]^T$. We train the reservoir using our measurement dataset and then predict forward for the duration of time it takes the system to traverse the interval of ρ that is of interest to us. We normalize each input to the reservoir, including the parameter

TABLE II. Hyperparameters for predicting the noiseless Lorenz system.

Reservoir size	N	1000
Average number of connections per node	$\langle d \rangle$	3
Spectral radius of the reservoir's adjacency matrix	ρ_s	0.85
Maximal coupling strength of the input to the reservoir	χ	0.6775
Tikhonov regularization parameter	α	2.15×10^{-11}
Numerical integration time step = RC time step	Δt	0.01

ρ , by dividing it by its root-mean-square value over the training interval so that we do not introduce a large bias into the activation function of the reservoir nodes. Tables II and III list the choice of RC hyperparameters that we use for all prediction tasks in this paper involving the noiseless (noisy) Lorenz system. We then use the sequence $\{z_m\}$ of resulting z maxima to display the results. Specifically, we plot the difference between consecutive z maxima, $z_{m+1} - z_m$, vs ρ in Fig. 12(a) and 12(b) for the trained reservoir (red dots) and the true orbit (black dots). We see that the reservoir was able to accurately capture all the transitions (for increasing ρ) discussed above, including the period halving cascade and the abrupt transition to chaos near $\rho \approx 169$. To more quantitatively verify that the reservoir predicts trajectories that mimic the long-term dynamical properties of the true system, we compare the cumulative distributions of z maxima for chosen intervals $[\rho^*, \rho^* + \Delta\rho]$, where $\Delta\rho = 0.5$, obtained from an ensemble of 2000 randomly initiated trajectories for both the true system, Eqs. (8), and the corresponding reservoir-predictions [see Figs. 12(c) and 12(d)]. Figure 12(d) shows a plot of $\Gamma(\rho)$ for different values of ρ in the prediction window showing that this quantity is small throughout most of the interval. There is, however, a large spike near $\rho = 167$, which corresponds to the transition from a period two orbit to a chaotic orbit. This spike occurs because, as was the case for the logistic map, the reservoir does not always get the exact location of the transition correct. However, the predicted transition point is still fairly close to the actual transition point. This can be seen by the immediate drop in $\Gamma(\rho)$ after the spike [and from Figs. 12(a) and 12(b)]. In Fig. 12(c), we show the cumulative probability distributions for $\rho_1^* = 139$,

TABLE III. Hyperparameters for predicting the noisy Lorenz system.

Reservoir size	N	1000
Average number of connections per node	$\langle d \rangle$	4
Spectral radius of the reservoir's adjacency matrix	ρ_s	0.85
Maximal coupling strength of the input to the reservoir	χ	0.4
Tikhonov regularization parameter	α	1.29×10^{-10}
Numerical integration time step = RC time step	Δt	0.01

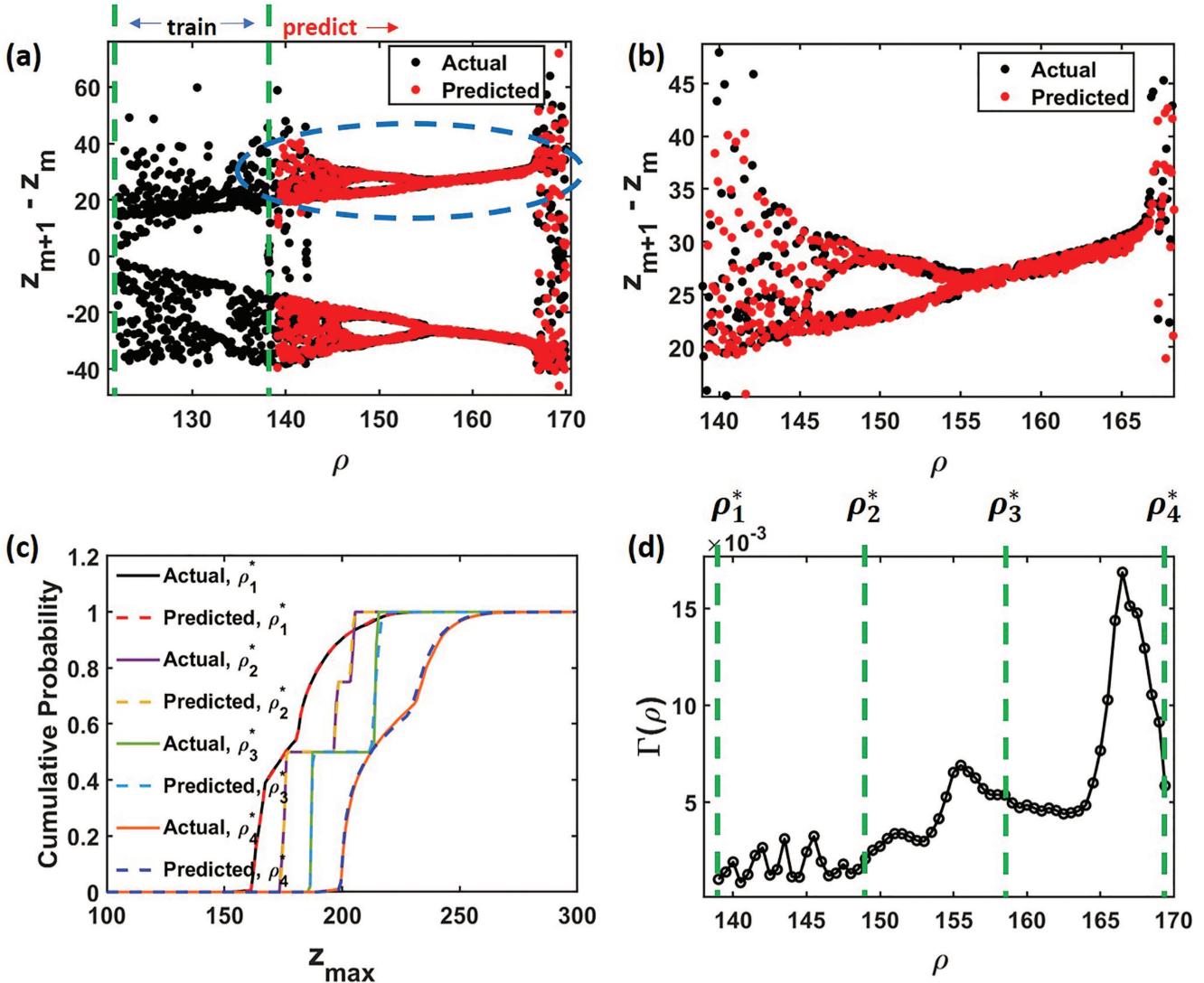


FIG. 12. (a) $(z_{m+1} - z_m)$ vs ρ for the true system (black dots) compared with that predicted by the reservoir (red dots). (b) Zoomed-in view of the region enclosed by the dashed blue oval in (a). (c) Cumulative probability distributions of z maxima for the true trajectories (solid curves) and the reservoir-predicted trajectories (dashed curves) for $\rho_1^* = 139$, $\rho_2^* = 149$, $\rho_3^* = 159$, $\rho_4^* = 169.4$, and $\Delta\rho = 0.5$. (d) $\Gamma(\rho)$ over the prediction interval. The vertical dashed lines indicate the points at which the cumulative probability distributions are shown in (c).

$\rho_1^* = 149$, $\rho_2^* = 159$, and $\rho_4^* = 169.4$. The corresponding errors in the distributions were $\Gamma(\rho_1^*) = 0.001$, $\Gamma(\rho_2^*) = 0.0021$, $\Gamma(\rho_3^*) = 0.0050$, and $\Gamma(\rho_4^*) = 0.0058$. These small values confirm that the reservoir has accurately replicated the non-stationary climate of the system at these ρ values.

Next, we demonstrate the reservoir's ability to predict the transitions in the opposite direction (i.e., for ρ decreasing in time); specifically, we take $\rho(t) = \rho_0 + \gamma t$ with $\gamma = -0.1$ and ρ_0 above the window. In addition, we also add white noise $\varepsilon_z(t)$ to the right-hand side of the Lorenz equation for $z(t)$, Eq. (8c), implemented by adding a random number sampled from a uniform distribution

over the interval $[-0.05, 0.05]$ at each $\Delta t = 0.01$ time step. We use the method developed in Sec. III to approximate a distribution $\tilde{\zeta}(\tilde{\varepsilon}_z)$ from which we can sample $\tilde{\varepsilon}_z$ and re-inject approximate noise [Fig. 9(b)] into our prediction scheme to replicate the effects of the dynamical noise present in the actual system. In Fig. 13, we show the results of the reservoir prediction of this noisy non-stationary process. Figure 13(a) and its zoomed-in view, Fig. 13(b), compare a plot of $z_{m+1} - z_m$ vs ρ for a true trajectory and for that predicted by the reservoir. We see that the reservoir is able to anticipate the intermittency-type regime transition from chaos to a period two orbit at $\rho \approx 165$, the noisy period doubling starting

near $\rho \approx 154$, and the noisy pairwise band-merging near $\rho \approx 145$ (as well as the noise-induced intermittency-like behavior just prior to the band-merging). Figure 13(d) shows $\Gamma(\rho)$ for various values of ρ in the prediction window. Again, we see a spike near $\rho \approx 165$, which is due to the reservoir's slightly inaccurate prediction of when the intermittency transition initiating the start of the window occurs. There is also a growing discrepancy between the reservoir predicted cumulative probability distribution and that obtained from actual trajectories toward the end of the prediction window due to error in the reservoir prediction of when the specific period doubling bifurcations and transition to chaos occur. Even so, although the predictions of these transitions are shifted from their true locations, their existence and occurrence are still well-anticipated. Figure 13(c) compares the approximate cumulative distributions at locations $\rho_1^* = 139$, $\rho_2^* = 149$, $\rho_3^* = 159$, and $\rho_4^* = 168.5$ as obtained from the true trajectories and as predicted by the reservoir for 2000 randomly initialized trajectories. The calculated errors at these points are small, $\Gamma(\rho_1^*) = 0.0123$, $\Gamma(\rho_2^*) = 0.0077$, $\Gamma(\rho_3^*) = 0.0025$, and $\Gamma(\rho_4^*) = 0.0008$, once again indicating that the reservoir has satisfactorily replicated the dynamics of the different regimes.

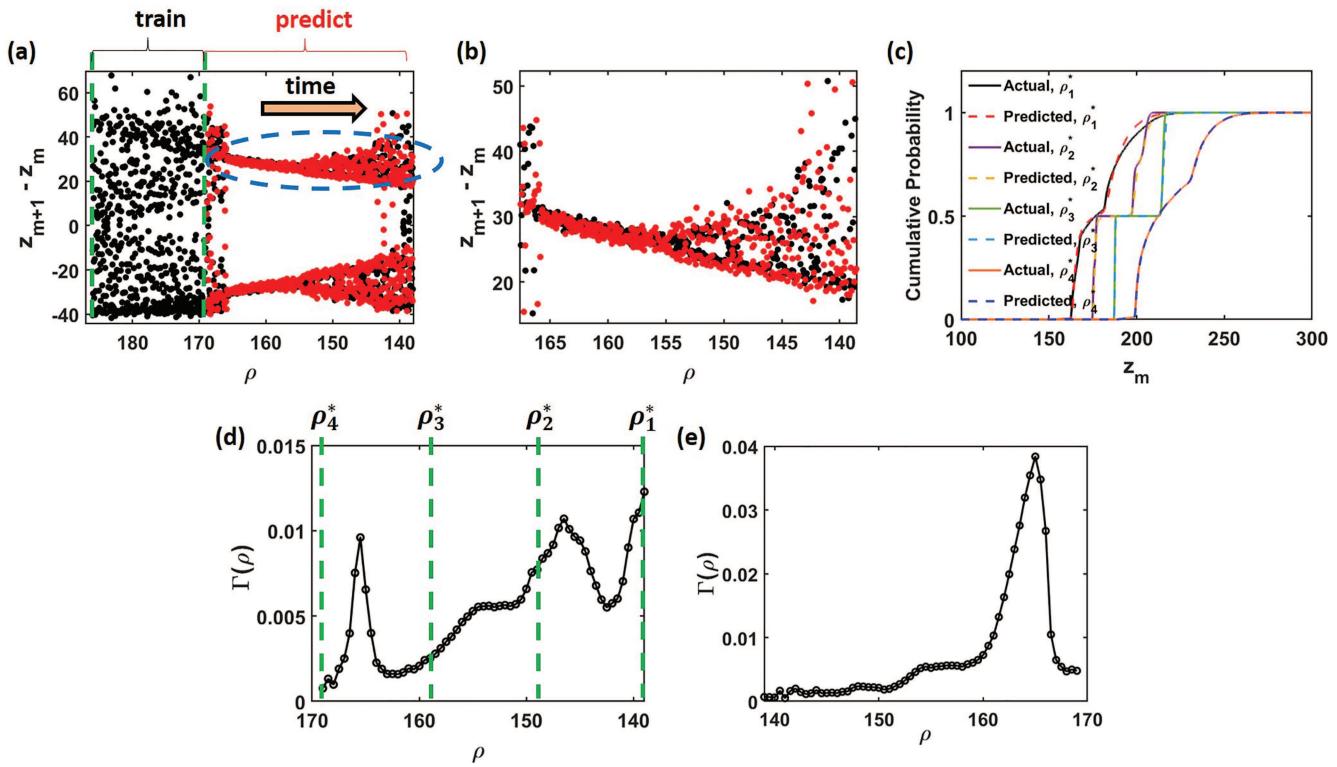


FIG. 13. (a) $(z_{m+1} - z_m)$ vs ρ from a numerical solution of the Lorenz equations (black dots) compared with the reservoir prediction (red dots that overlay the black dots). (b) Zoomed-in view of the region enclosed by the dashed blue oval in (a). (c) Cumulative probability distributions of z_m from an ensemble of true (solid curves) and reservoir-predicted trajectories (dashed curves) for $\rho_1^* = 139$, $\rho_2^* = 149$, $\rho_3^* = 159$, $\rho_4^* = 168.5$, and $\Delta\rho = 0.5$. (d) $\Gamma(\rho)$ vs ρ over the prediction interval [for case with $\rho(t)$ decreasing in time]. The vertical dashed lines indicate the ρ values for which the cumulative probability distributions are shown in (c). (e) $\Gamma(\rho)$ for the case of the same noisy Lorenz system but with $\rho(t)$ increasing with time ($\gamma = 0.1$).

For comparison, Fig. 13(e) shows $\Gamma(\rho)$ for the same noisy Lorenz system but with ρ_0 under the window and $\rho(t)$ increasing with time; i.e., $\gamma = +0.1$. We note that this plot does not look similar to the previous case [Fig. 13(d)] with $\rho(t)$ decreasing in that there is no spike in Γ near the start of the prediction window. We do, however, see that there is a relatively larger error in the reservoir's prediction of the transition from a period 2 orbit to chaos.

We now revisit the case [Eq. (6) and Fig. 7 from Sec. III] for the noiseless Lorenz system, but for a case in which the parameter variation is non-monotonic,

$$\rho(t) = \rho_0 + \rho_1 \sin(\theta) + \mu t, \quad (9a)$$

$$\theta(t) = \gamma t, \quad (9b)$$

where we set $\rho_0 = 120$, $\rho_1 = 10$, $\gamma = 0.02$, and $\mu = 0.1$. Figure 14(a) shows the time variation of $\rho(t)$. Note that the training is over an interval of time during which $\rho(t)$ is approximately linear and increasing. The prediction is then carried out as $\rho(t)$ changes highly nonlinearly and, furthermore, contains an interval where $\rho(t)$ decreases with time, $d\rho/dt < 0$. Figure 14(b), and the blow-up in Fig. 14(c), compares $z_{m+1} - z_m$ vs time for a true trajectory

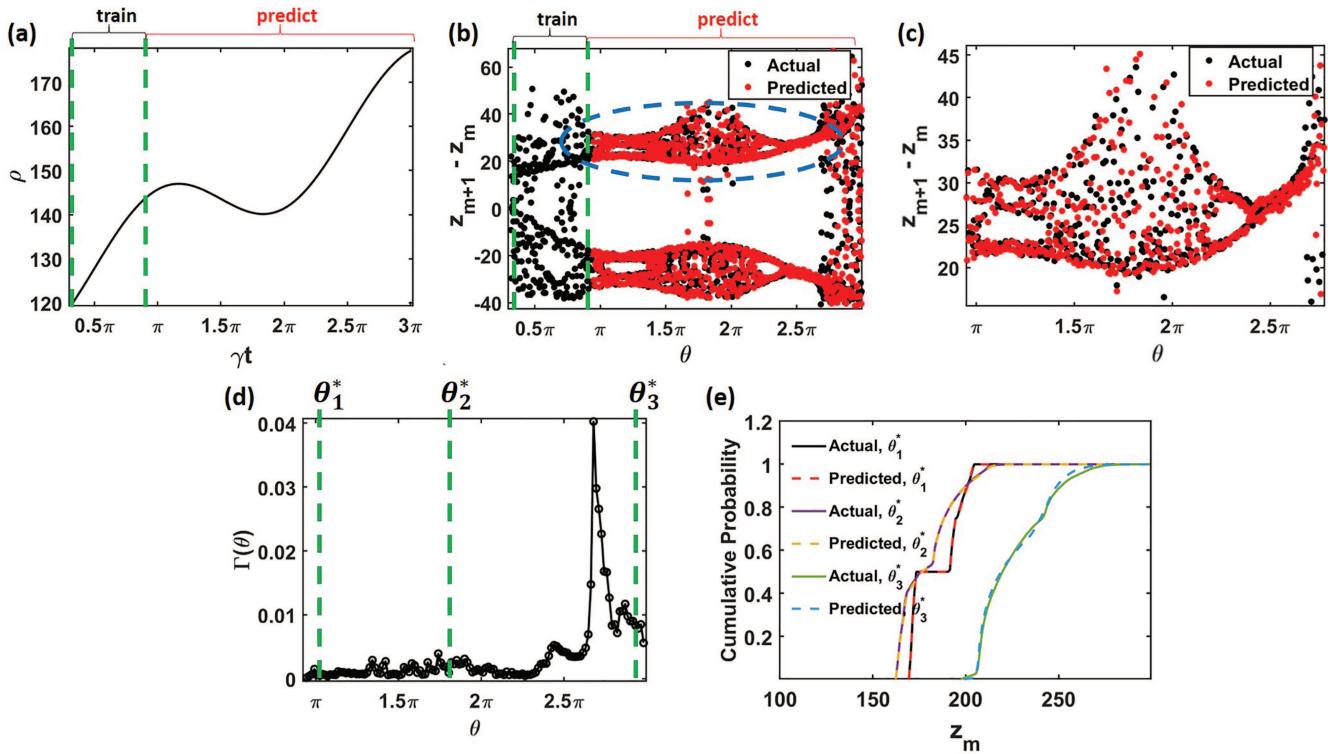


FIG. 14. (a) $\rho(t)$ as a function of t . (b) $(z_{m+1} - z_m)$ vs γt for the true trajectory (black dots) and for the reservoir prediction (red dots that overlay the black dots). (c) A zoomed-in view of the oval region indicated in (b). (d) $\Gamma(\theta)$ over the prediction interval. The vertical dashed lines indicate the points at which the cumulative probability distributions are shown in (e). (e) Cumulative distributions of the map obtained from the real trajectory (solid curves) and from the reservoir-predicted trajectory (dashed curves) at $\theta_1^* = \pi$, $\theta_2^* = 1.8\pi$, and $\theta_3^* = 2.9\pi$ with $\Delta\theta = 0.05$.

evolving according to the Lorenz system with that from the reservoir prediction. Figure 14(d) shows the values of $\Gamma(\theta)$ over the prediction interval. As with the previous cases, we see a spike near the intermittency transition from the period 2 orbit to chaos near $\theta \approx 2.67\pi$ (which corresponds to $\rho \approx 165$), which is due to the reservoir's slightly inaccurate prediction of when the transition occurs. The approximate cumulative distributions computed over small intervals $\Delta\theta = 0.05$ are obtained at three times $\theta_1^* = \pi$, $\theta_2^* = 1.8\pi$, and $\theta_3^* = 2.9\pi$ [shown in Fig. 14(e)]. At these θ values, the calculated errors between the distributions obtained from the actual orbits and those obtained from the reservoir predictions are small, $\Gamma(\theta_1^*) = 0.0006$, $\Gamma(\theta_2^*) = 0.0014$, and $\Gamma(\theta_3^*) = 0.0084$. We note that, although the reservoir was trained over an interval during which ρ increased approximately linearly, it was still able to predict the transitions and climate as ρ changed nonlinearly. In particular, the reservoir was able to accurately capture the transition from the two-piece chaotic set to a single-band chaotic set near $\theta \approx 1.75\pi$. After this point, ρ increases rapidly, leading to a transition from a period 2 orbit to chaos, which the reservoir is able to accurately predict. To summarize, our results for the parameter time variation, Eq. (9), show that the reservoir was able to replicate the long-term dynamical properties of the non-stationary Lorenz system when it was trained on a set of measurements

where the driving parameter ρ varied differently than it did during prediction.

VI. APPLICATION OF RC TO PREDICTING THE KURAMOTO-SIVASHINSKY SYSTEM

We consider a scalar field $w(x, t)$ in one spatial variable x and time t , with periodic boundary conditions $w(x, t) = w(x + 2\pi, t)$ and with $w(x, t)$ satisfying the noisy, possibly non-stationary, Kuramoto-Sivashinsky (KS) equation,

$$\frac{\partial w(x, t)}{\partial t} + 2w(x, t)\frac{\partial w(x, t)}{\partial x} + \frac{\partial^2 w(x, t)}{\partial x^2} + \kappa \frac{\partial^4 w(x, t)}{\partial x^4} = \xi(x, t), \quad (10)$$

where $\xi(x, t)$ represents spatially dependent noise that is white in time and the parameter κ can be either constant in time, in which case, we write $\kappa = \kappa_0$, or varying with time [$\kappa = \kappa(t)$]. Furthermore, for simplicity, following Christiansen *et al.*,⁹¹ we consider solutions $w(x, t)$ having odd symmetry about $x = 0$, $w(x, t) = -w(-x, t)$, in which case we can expand $w(x, t)$ in a Fourier series of the form

$$w(x, t) = i \sum_{p=-\infty}^{+\infty} a_p(t) e^{ipx}, \quad (11)$$

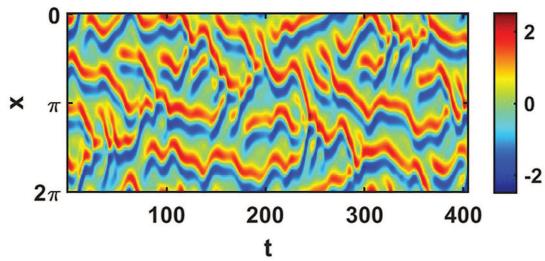


FIG. 15. A solution, $w(x, t)$, of the KS equation for $\kappa_0 = 0.02975$. The vertical axis shows x , the horizontal shows t (in seconds), and the values of $w(x, t)$ are color-coded.

where $a_p(t) = -a_{-p}(t)$, $a_p(t)$ is real, and $a_0(t) = 0$. (In the general, non-symmetric case, the Fourier coefficients, rather than being purely imaginary, ia_p , would have real and imaginary parts.) In order that the noise not break this convenient symmetry, we take the noise to also have the same symmetry, in which case its Fourier representation has the same form as Eq. (11),

$$\xi(x, t) = i \sum_{p=-\infty}^{+\infty} \xi_p(t) e^{ipx}. \quad (12)$$

Hence, $\xi_p(t)$ are real white noise terms that are uncorrelated for $p \neq p'$ and delta-correlated in time, $\langle \xi_p(t) \xi_{p'}(t') \rangle = (\text{constant}) \times \delta(t - t')$, where the angle brackets denote an ensemble average.

Substituting (11) and (12) into (10), we obtain an infinite, coupled set of first order ordinary differential equations,⁹¹

$$\begin{aligned} da_p(t)/dt &= (p^2 - \kappa p^4)a_p(t) \\ &- p \sum_{q=-\infty}^{\infty} a_q(t)a_{p-q}(t) + \xi_p(t). \end{aligned} \quad (13)$$

We solve Eq. (13) numerically using the Runge–Kutta fourth order method with a time step of 0.0001 and using a truncation approximation at $p_{\max} = 16$ (i.e., we set $a_p = 0$ for $p > p_{\max}$). Figure 15 shows an example of a stationary system solution for $w(x, t)$ and $\kappa_0 = 0.02975$, where time is plotted horizontally, x is plotted vertically, and the value of w is color-coded.

Considering the noiseless [$\xi(x, t) = 0$], stationary ($\kappa = \kappa_0$) version of Eq. (10), again following Ref. 91, we use a Poincaré map with surface of section, $a_1 = 0$, to obtain a stationary system bifurcation diagram showing the various regime transitions and bifurcations. This is shown in Fig. 16(a), which plots the values of $a_6(t_n)$ at many surface of section crossings ($t = t_n$) vs κ_0 for κ_0 in the range $0.0296 < \kappa_0 < 0.02989$. In this range of κ , we see chaotic dynamics, interspersed with periodic windows of different periodicities (e.g., the period 5 window near $\kappa_0 \approx 0.02985$). We also see that the size of the attractor is larger for smaller values of κ_0 .

Applying our RC prediction method for a non-stationary, noiseless case,

$$\kappa(t) = \kappa(0) + \gamma t, \quad (14)$$

with parameters given in Table IV, we obtain the results in Fig. 16(b), comparing the RC prediction (red dots) with the

numerical solution of the KS equation (black dots). Here, we set $\gamma = 5.5 \times 10^{-8}$ and $\kappa(0) = 0.029825$. In this case, the (relatively) fast drift of the parameter κ results in the loss of fine detail in the regime transitions in this period 5 window (e.g., we cannot discern a period doubling cascade in each of the five branches near the lower κ range of the window). Alternatively, we see a sudden transition from chaos below $\kappa \approx 0.029844$ to a period 5 orbit above it and similarly an abrupt transition from the period 5 orbit below $\kappa \approx 0.029854$ to chaos above it. We see from Fig. 16(b) that the reservoir system is able to capture the transitions from chaos to periodic and then back to chaos. The red dots represent the reservoir-predicted surface of section crossings of the a_6 -coordinate, $a_6(t_n)$, and the black dots represent $a_6(t_n)$ obtained from a numerical solution of Eq. (13). Figure 16(c) shows a plot of $\Gamma(\kappa)$ over this period 5 prediction window, calculated using an ensemble of 1000 randomly selected initial conditions and for $\Delta\kappa = 5 \times 10^{-7}$. As before, we see that the value of Γ is small throughout the window, indicating that the reservoir-predicted trajectories adequately replicate the non-stationary climate of this system. The spikes in Γ near $\kappa \approx 0.029844$ and $\kappa \approx 0.029853$ are due to the reservoir's slightly inaccurate prediction of the exact location of the transition from chaotic to periodic motion and from periodic back to chaotic, respectively. The narrow width of these spikes indicates that this error is small. Figure 16(d) shows the cumulative probability distributions obtained from the numerically integrated solutions of Eq. (13) (solid curves) and from the reservoir-predicted trajectories (dashed curves) at $\kappa_1^* = 0.0298400$, $\kappa_2^* = 0.0298500$, and $\kappa_3^* = 0.0298565$. The corresponding values of Γ are $\Gamma(\kappa_1^*) = 0.0026$, $\Gamma(\kappa_2^*) = 0.0038$, and $\Gamma(\kappa_3^*) = 0.0038$. The agreement between the curves and the small values of Γ shows that the reservoir-predicted trajectories accurately predict the dynamics of the different regimes.

To ensure reservoir stability during the closed-loop prediction phase for the non-stationary, noiseless case, we found it necessary to insert a small noise $\delta u(t)$, sampled from a uniform distribution over the interval $[-\delta u, \delta u]$, into the reservoir update equation [Eq. (1)] at each time step,

$$r(t + \Delta t) = \tanh(Ar(t) + W_{in}(u(t) + \delta u(t))). \quad (15)$$

For the above example, we used $\delta u = 9 \times 10^{-5}$. In addition, we used a “multi-pass” training scheme to obtain the above results as follows. The reservoir is run in the open-loop training configuration using Eq. (15), the training data $\{u(t), \kappa(t)\}_{t=-T}^{t_0}$, and a realization of the input noise process $\{\delta u_1(t)\}_{t=-T}^{t_0}$. The reservoir states $\{r(t)\}$ are collected. The reservoir, after re-initializing it to some starting value $r(0)$, is again run in the open-loop configuration using Eq. (15), the same training data $\{u(t), \kappa(t)\}_{t=-T}^{t_0}$, and a *different* realization of the input noise process $\{\delta u_2(t)\}_{t=-T}^{t_0}$. The reservoir states are collected again. This procedure is repeated a predetermined number of times (ten in this case), and a linear regression is performed on the collection of all resulting reservoir states to determine the output weight matrix W_{out} , which maps the collection of reservoir states $\{r(t)\}$ to $\{u(t)\}_{t=-T}^{t_0}$. (The closed-loop prediction phase is performed in the usual way, i.e., without adding any noise to the reservoir update equation.) This method uses a single set of training data and multiple realizations of an input noise process to allow for determination of a more robust W_{out} . (A detailed exposition and analysis of noise

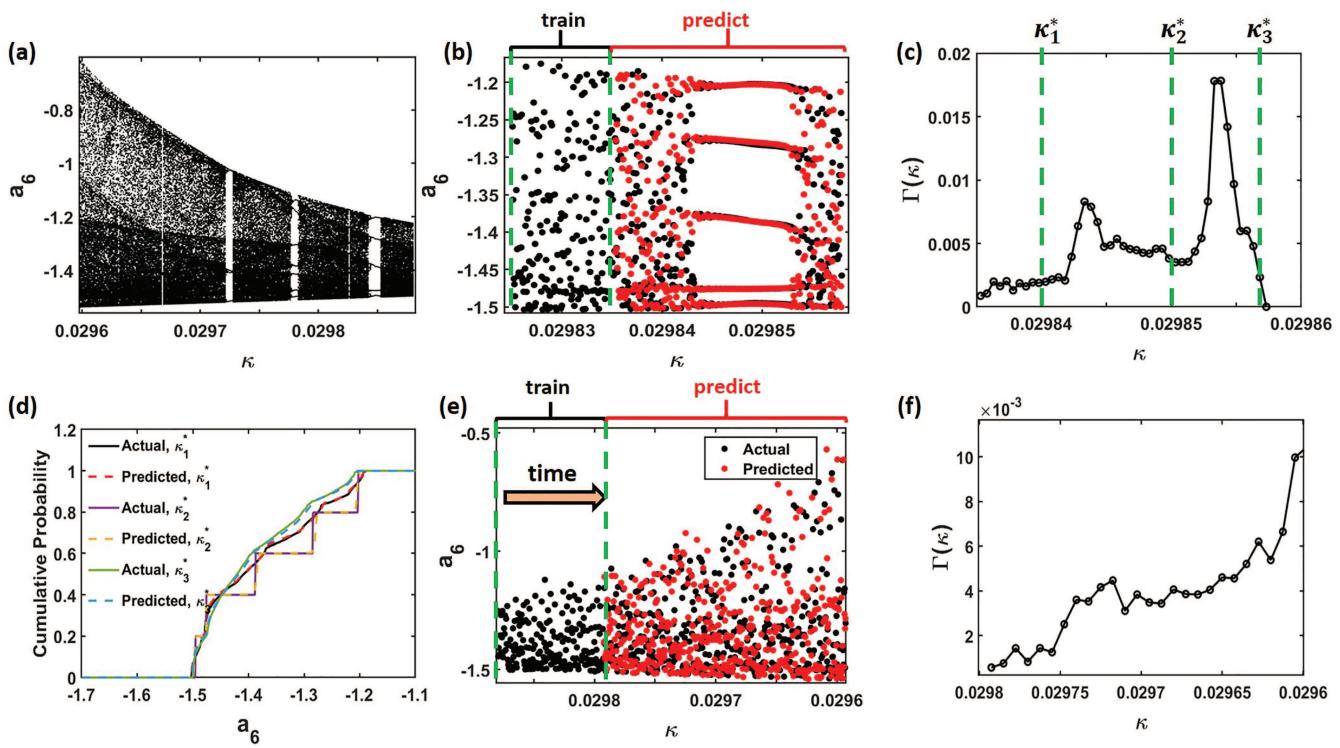


FIG. 16. (a) Bifurcation diagram of the stationary Kuramoto–Sivashinsky system [Eq. (10)]. (b) Comparison of the actual $a_6(t_n)$ (black dots) for a numerically integrated trajectory of Eq. (13) and the corresponding reservoir-predicted $a_6(t_n)$ (red dots overlaying the black dots) for a linear drift in the parameter κ . (c) $\Gamma(\kappa)$ over the period 5 prediction window. The dashed vertical green lines indicate the points at which the cumulative probability distributions are shown in (d). (d) Cumulative probability distributions of the map obtained from the numerical solutions of Eq. (13) (solid curves) and from the reservoir-predicted trajectories (dashed curves) at $\kappa_1^* = 0.029\,840\,0$, $\kappa_2^* = 0.029\,850\,0$, $\kappa_3^* = 0.029\,856\,5$, and $\Delta\kappa = 5 \times 10^{-7}$. (e) $a_6(t_n)$ vs κ for a numerically determined trajectory of Eq. (13) (black dots) and that predicted by the trained reservoir (red dots). (f) $\Gamma(\kappa)$ over the prediction window for the case in (e).

stabilization of machine learning prediction will be the subject of a future publication.)

Next, we apply our RC prediction method to the noisy, non-stationary case with $\gamma = -5 \times 10^{-7}$, and, as in Sec. V, the noise (ξ_p) is numerically implemented by adding a random number at each

TABLE IV. Hyperparameters for predicting the noiseless Kuramoto–Sivashinsky system.

Reservoir size	N	2000
Average number of connections per node	$\langle d \rangle$	3
Spectral radius of the reservoir's adjacency matrix	ρ_s	0.45
Maximal coupling strength of the input to the reservoir	χ	0.6775
Tikhonov regularization parameter	α	1.67×10^{-7}
Numerical integration time step for Eq. (13)	Δt	0.0001
RC time step	Δt_{RC}	0.005

iteration of our numerical solution of Eq. (13) for a_p , $0 < p \leq p_{\max}$, where the random number, independent for each value of p , is uniformly chosen in the range $[-5 \times 10^{-5}, 5 \times 10^{-5}]$. Table V lists the hyperparameters used for this task. As with the previous case, we use the multi-pass training scheme with ten passes (i.e., we run the reservoir in the open-loop configuration ten times using the same

TABLE V. Hyperparameters for predicting the noisy Kuramoto–Sivashinsky system.

Reservoir size	N	2000
Average number of connections per node	$\langle d \rangle$	2
Spectral radius of the reservoir's adjacency matrix	ρ_s	0.65
Maximal coupling strength of the input to the reservoir	χ	0.9
Tikhonov regularization parameter	α	7.74×10^{-10}
Numerical integration time step for Eq. (13)	Δt	0.0001
RC time step	Δt_{RC}	0.005

set of training data and a different noise realization each time) and the noise sampled from the same distribution as before. We consider the range $\kappa \in [0.02960, 0.02988]$, in which, due to the rate of non-stationarity and added dynamical noise, we cannot discern the periodic windows [see the black dots in Fig. 16(e) representing the $a_6(t_n)$ obtained from a numerical solution of Eq. (13)], which are visible in the stationary bifurcation diagram of Fig. 16(a). Here, we see that the size of the attractor gradually increases with decreasing κ . The reservoir-predicted trajectory [represented by the red dots in Fig. 16(e)] accurately captures this gradual increase in the size of the attractor. Figure 16(f) shows $\Gamma(\kappa)$, obtained from 5000 randomly initialized trajectories, over the prediction window. From the small value of Γ over the prediction interval, we see that the reservoir-predicted trajectory also accurately captures the κ -dependent density variation of $a_6(t_n)$. This indicates that the non-stationary climate of this system is well-replicated by the reservoir-predicted trajectories.

VII. CONCLUSION

In this paper, we have formulated and tested a machine learning method for prediction of the behavior or “climate” (defined in Sec. I), of non-stationary dynamical systems (both noiseless and noisy) from time series data of the state of the system and knowledge of the time dependence of a system parameter responsible for the non-stationary but with no further knowledge of the nature of the system to be predicted. The methods introduced in Sec. IV are based on a previously formulated machine learning feedback prediction scheme for the dynamics of stationary systems (Sec. II), which we modify in order to adapt it for the purpose of climate prediction of noiseless non-stationary dynamical systems (Sec. IV A) and of noisy non-stationary dynamical systems (Sec. IV B). In the case of prediction for noiseless non-stationary systems, the modified prediction method (Fig. 5) is the stationary system prediction method (Fig. 1) altered to include input of the known time-dependent parameter {and this input applies for both the training phase [Fig. 5(a)] and the prediction phase [Fig. 5(b)]}. In the case of noisy non-stationary systems, the prediction method (Fig. 9) contains a further alteration in which dynamical noise approximating that of the unknown system to be predicted is injected at the output side of the machine learning system; we describe how to obtain from the available time series data the required approximate probability distribution of the actual noise.

These methods are tested on non-stationary versions of three systems: a simple discrete-time map (the logistic map, Sec. IV), a continuous time set of three autonomous ordinary differential equations (the Lorenz 63 equations, Sec. V), and a spatiotemporal partial differential equation (the Kuramoto–Sivashinsky equation, Sec. VI). Evaluations for the resulting tests are made through comparison of orbit characteristics of the actual systems with those predicted using plots of orbits, cumulative probability distributions, and the Wasserstein metric for the distance between probability distributions.

Some dynamical orbit features whose predictions are tested include the following:

- Apparently continuous evolution of the distribution of chaotic orbit densities in chaotic regions without apparent regime transitions.
- Non-stationarity induced phenomenon, such as the apparent delay and increased abruptness of period doublings [Fig. 3(b)].
- The effects of dynamical noise, e.g., the apparent blurring of features present in stationary system bifurcation diagrams and intermittent bursting out of the small multiple piece chaotic attractors preceding interior crisis transitions [Fig. 8 and 10(b)].
- Various types of regime transitions, including the forward and backward versions of period doubling cascades of periodic orbits followed by pairwise splitting cascades of chaotic attractors, intermittency transitions between periodicity and chaos, and crisis transitions.

Results of our tests show that the overall dependence of the future pattern of climate variation with time is extremely well-predicted and that the quantitative agreement between the predicted and actual behavior is also remarkably good, although there are some notable small to moderate shifts in the occurrences of regime transitions. Based on these results, we expect that machine learning will be of great future utility as a means for predicting the statistical behavior (climate) of unknown, non-stationary dynamical systems from time series data. Resolution of the question of how, and whether, this expectation is realized for many potential practical applications (see Sec. I) awaits further study.

ACKNOWLEDGMENTS

We thank Brian Hunt for insightful comments, suggestions, and advice. The authors acknowledge the University of Maryland supercomputing resources (<http://hpc.umd.edu>) made available for conducting the research reported in this paper. This work was supported, in part, by the National Science Foundation (NSF) under Grant Nos. DMS-1813027 and DGE-1632976 and by the Defense Advanced Research Projects Agency (DARPA) grant (No. W31P4Q-20-C-0077).

APPENDIX: ANALYSIS OF THE NOISE-INJECTION SCHEME

As shown in Fig. 9, to approximately replicate the effects of uncorrelated stochastic forcing, we inject an uncorrelated noise term $\tilde{\varepsilon}_n$ into the reservoir output at each time step. We desire that the distribution $\tilde{\zeta}$ of the injected noise $\tilde{\varepsilon}_n$ is a reasonable approximation to the distribution ζ of the actual noise ε_n . To obtain this estimate, during the training, we take the difference between the reservoir output \tilde{v}_{n+1} and the actual noisy system trajectory v_{n+1} as our noise estimate, $\tilde{\varepsilon}_n = v_{n+1} - \tilde{v}_{n+1}$. If the reservoir output \tilde{v}_{n+1} were the actual noiseless output v_{n+1} , then $\tilde{\varepsilon}_n = \varepsilon_n$, but the reservoir makes some error $\Delta v_{n+1} = \tilde{v}_{n+1} - v_{n+1}$ and $\tilde{\varepsilon}_n$ is only an approximation of the true noise; i.e., $\tilde{\varepsilon}_n = \varepsilon_n + \Delta v_{n+1}$. Thus, to the extent that the reservoir error Δv_{n+1} is small compared to the actual noise ε_n , our approximation will be good. Noting that ε_n is an uncorrelated noise, while both v_{n+1} and \tilde{v}_{n+1} come from deterministic mappings of a state x_n that is independent of ε_n (where these mappings are the mapping of the system to be predicted and its reservoir approximation), we see that the reservoir error Δv_{n+1} is uncorrelated with

ε_n . Letting $\zeta_\Delta(\Delta v)$ denote the distribution function of Δv computed from the time series Δv_n , by virtue of $\tilde{\varepsilon}_n = \varepsilon_{n+1} + \Delta v_{n+1}$, $\tilde{\zeta}$ is given by the convolution

$$\tilde{\zeta}(\tilde{\varepsilon}) = \int \zeta(\varepsilon) \zeta_\Delta(\tilde{\varepsilon} - \varepsilon) d\varepsilon.$$

Roughly speaking, the approximation $\tilde{\zeta}(\tilde{\varepsilon})$ is obtained by broadening $\zeta(\varepsilon)$ by the width of ζ_Δ , which, if small compared to the width of ζ , may be expected to yield a good approximation of ζ . Using a flat-top distribution for $\zeta(\varepsilon)$ (as in our previous numerical experiments), we demonstrate this for the non-stationary logistic map for two noise levels [widths of $\zeta(\varepsilon)$], 0.0005, 0.002, and $\gamma = 10^{-5}$, as follows. An ensemble of randomly chosen initial conditions is evolved by the actual orbit x_n , obtained using the logistic map, with noise ε_n , and the trained reservoir is used to obtain the

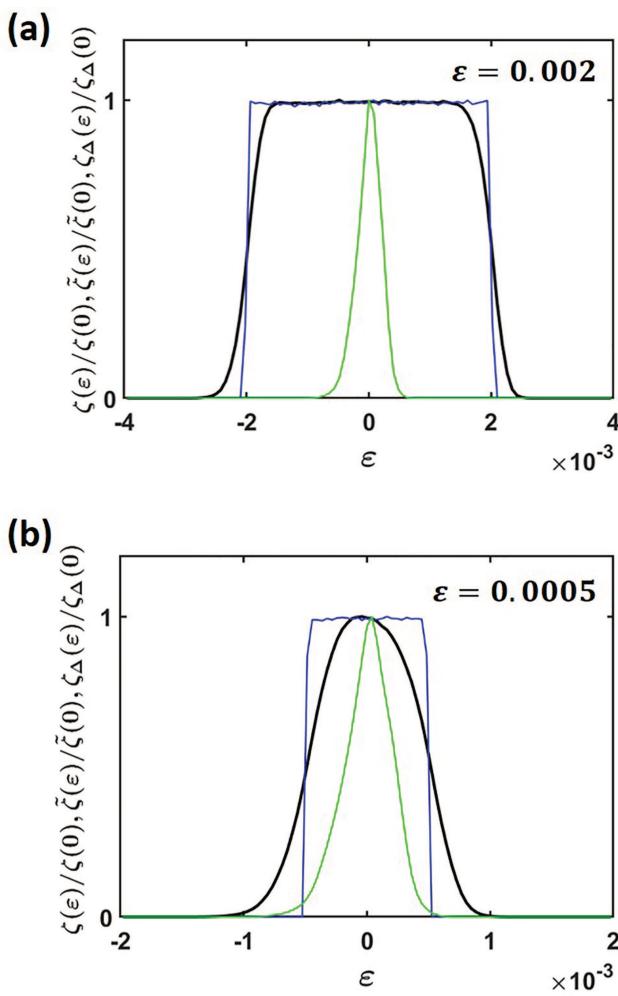


FIG. 17. (a) and (b) show comparisons of $\tilde{\zeta}$ (black lines), ζ (blue line), and ζ_Δ (green line) for different noise levels [widths of $\zeta(\varepsilon)$], 0.002, and 0.0005, respectively.

time series, $\tilde{\varepsilon}_n = x_{n+1} - \tilde{v}_{n+1}$, where \tilde{v}_{n+1} is the reservoir output when its input is x_n . In addition, we also use our determination of the actual orbit x_n to obtain the time series v_n via Eq. (7). We then use these time series to form histogram approximations to ζ , $\tilde{\zeta}$, and ζ_Δ . Figures 17(a) and 17(b) show numerical results for the distributions ζ_Δ (green lines), ζ (blue lines), and $\tilde{\zeta}$ (black lines). Figure 17(a) shows a logistic case where the dynamical noise level is substantially larger than the reservoir error (width of ζ_Δ), and $\tilde{\zeta}$ provides a good approximation to ζ . Figure 17(b) provides an example where the reservoir error is of the same order as the noise, leading $\tilde{\zeta}$ to noticeably deviate from ζ , although we note that the reservoir training error could be decreased by using a larger reservoir with well-optimized hyperparameters so that $\tilde{\zeta}$ better approximates ζ .

DATA AVAILABILITY

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

REFERENCES

- ¹H. Jaeger and H. Haas, “Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication,” *Science* **304**, 78 (2004).
- ²J. Pathak, B. R. 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).
- ³J. Pathak, A. Wikner, R. Fussel, S. Chandra, B. Hunt, M. Girvan, and E. Ott, “Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge based model,” *Chaos* **28**, 041101 (2018).
- ⁴J. 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).
- ⁵Z. Lu, B. R. Hunt, and E. Ott, “Attractor reconstruction by machine learning,” *Chaos* **28**, 061104 (2018).
- ⁶Z. Shi and M. Han, “Support vector echo-state machine for chaotic time series prediction,” *IEEE Trans. Neural Netw.* **18**, 359 (2007).
- ⁷S. Mukherjee, E. Osuna, and F. Girosi, “Nonlinear prediction of chaotic time series using support vector machines,” in *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop* (IEEE, 1997) pp. 511–520.
- ⁸Q. Ma, Q. Zheng, H. Peng, T. Zhong, and L. Xu, “Chaotic time series prediction based on evolving recurrent neural networks,” in *2007 International Conference on Machine Learning and Cybernetics* (IEEE, 2007), Vol. 6, pp. 3496–3500.
- ⁹R. Chandra, Y.-S. Ong, and C.-K. Goh, “Co-evolutionary multi-task learning with predictive recurrence for multi-step chaotic time series prediction,” *Neurocomputing* **243**, 21–34 (2017).
- ¹⁰D. Li, M. Han, and J. Wang, “Chaotic time series prediction based on a novel robust echo state network,” *IEEE Trans. Neural Netw. Learn. Syst.* **23**, 787–799 (2012).
- ¹¹M. Sangiorgio and F. Dercole, “Robustness of LSTM neural networks for multi-step forecasting of chaotic time series,” *Chaos, Solitons Fractals* **139**, 110045 (2020).
- ¹²Z. Tian, “Echo state network based on improved fruit fly optimization algorithm for chaotic time series prediction,” *J. Ambient Intell. Humaniz. Comput.* (published online).
- ¹³A. Vaughan and S. V. Bohac, “Real-time, adaptive machine learning for non-stationary, near chaotic gasoline engine combustion time series,” *Neural Netw.* **70**, 18–26 (2015).
- ¹⁴A. Griffith, A. Pomerance, and D. J. Gauthier, “Forecasting chaotic systems with very low connectivity reservoir computers,” *Chaos* **29**, 123108 (2019).
- ¹⁵G. Shen, J. Kurths, and Y. Yuan, “Sequence-to-sequence prediction of spatiotemporal systems,” *Chaos* **30**, 023102 (2020).

- ¹⁶R. Follmann and J. E. Rosa, "Predicting slow and fast neuronal dynamics with machine learning," *Chaos* **29**, 113119 (2019).
- ¹⁷S. Herzog, F. Wörgötter, and U. Parlitz, "Convolutional autoencoder and conditional random fields hybrid for predicting spatial-temporal chaos," *Chaos* **29**, 123116 (2019).
- ¹⁸A. Chattopadhyay, P. Hassanzadeh, and D. Subramanian, "Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: Reservoir computing, artificial neural network, and long short-term memory network," *Nonlinear Process. Geophys.* **27**, 373–389 (2020).
- ¹⁹D. Chen and W. Han, "Prediction of multivariate chaotic time series via radial basis function neural networks," *Complexity* **18**, 55 (2013).
- ²⁰W. Huang, Y. Li, and Y. Huang, "Deep hybrid neural network and improved differential neuroevolution for chaotic time series prediction," *IEEE Access* **8**, 159552–159565 (2020).
- ²¹H. Fan, J. Jiang, C. Zhang, X. Wang, and Y.-C. Lai, "Long-term prediction of chaotic systems with machine learning," *Phys. Rev. Res.* **2**, 012080 (2020).
- ²²A. Faqih, A. P. Lianto, and B. Kusumoputro, "Mackey-Glass chaotic time series prediction using modified RBF neural networks," in *Proceedings of the 2nd International Conference on Software Engineering and Information Management (ICSIM 2019)* (Association for Computing Machinery, New York, 2019), pp. 7–11.
- ²³P. Antonik, M. Gulina, J. Pauwels, and S. Massar, "Using a reservoir computer to learn chaotic attractors, with applications in chaos synchronization and cryptography," *Phys. Rev. E* **98**, 012215 (2018).
- ²⁴D. Nguyen, S. Ouala, L. Drumetz, and R. Fablet, "EM-like learning chaotic dynamics from noisy and partial observations," *arXiv:1903.10335* (2019).
- ²⁵F. Z. Xing, E. Cambria, and X. Zou, "Predicting evolving chaotic time series with fuzzy neural networks," in *2017 International Joint Conference on Neural Networks (IJCNN)* (IEEE Press, 2017) pp. 3176–3183.
- ²⁶J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys.* **57**, 617 (1985).
- ²⁷G. Drótos, T. Bódai, and T. Tél, "Quantifying nonergodicity in nonautonomous dissipative dynamical systems: An application to climate change," *Phys. Rev. E* **94**, 022214 (2016).
- ²⁸G. Drótos, T. Bódai, and T. Tél, "Probabilistic concepts in a changing climate: A snapshot attractor picture," *J. Clim.* **28**, 3275–3288 (15 Apr. 2015).
- ²⁹F. J. Romeiras, C. Grebogi, and E. Ott, "Multifractal properties of snapshot attractors of random maps," *Phys. Rev. A* **41**, 784–799 (1990).
- ³⁰M. D. Chekroun, E. Simonnet, and M. Ghil, "Stochastic climate dynamics: Random attractors and time-dependent invariant measures," *Phys. D* **240**, 1685–1700 (2011).
- ³¹S. H. Lim, L. Theo Giorgini, W. Moon, and J. S. Wettlaufer, "Predicting critical transitions in multiscale dynamical systems using reservoir computing," *Chaos* **30**, 123126 (2020).
- ³²M. J. Feigenbaum, "Quantitative universality for a class of nonlinear transformations," *J. Stat. Phys.* **19**, 25–52 (1978).
- ³³M. J. Feigenbaum, "Universal behavior in nonlinear systems," *Phys. D* **7**, 16–39 (1983).
- ³⁴Y. Pomeau and P. M. Manneville, "Intermittent transition to turbulence in dissipative dynamical systems," *Commun. Math. Phys.* **74**, 189–197 (1980).
- ³⁵C. Grebogi, E. Ott, and J. A. Yorke, "Crises, sudden changes in chaotic attractors, and transient chaos," *Phys. D* **7**, 181–200 (1983).
- ³⁶C. Grebogi, E. Ott, and J. A. Yorke, "Critical exponents of chaotic transients in nonlinear dynamical systems," *Phys. Rev. Lett.* **57**, 1284 (1986).
- ³⁷L.-W. Kong, H.-W. Fan, C. Grebogi, and Y.-C. Lai, "Machine learning prediction of critical transition and system collapse," *arXiv:2012.01545* (2019).
- ³⁸T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber, "Tipping elements in the Earth's climate system," *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1786–1793 (2008).
- ³⁹P. Ashwin and A. S. von der Heydt, "Extreme sensitivity and climate tipping points," *J. Stat. Phys.* **179**, 1531–1552 (2020).
- ⁴⁰S. O. Rasmussen, M. Bigler, S. P. Blockley, T. Blunier, S. L. Buchardt, H. B. Clausen, I. Cvijanovic, D. Dahl-Jensen, S. J. Johnsen, H. Fischer, V. Gkinis, M. Guillet, W. Z. Hoek, J. J. Lowe, J. B. Pedro, T. Popp, I. K. Seierstad, J. P. Steffensen, A. M. Svensson, P. Valerenga, B. M. Vinther, M. J. Walker, J. J. Wheatley, and M. Winstrup, "A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: Refining and extending the INTIMATE event stratigraphy," *Quat. Sci. Rev.* **106**, 14–28 (2014).
- ⁴¹R. T. Pierrehumbert, "Warming the world," *Nature* **432**, 677 (2004).
- ⁴²A. Zhang, G. Knorr, G. Lohmann, and S. Barker, "Abrupt North Atlantic circulation changes in response to gradual CO₂ forcing in a glacial climate state," *Nat. Geosci.* **10**, 518–523 (2017).
- ⁴³Q. Schiermeier, "A sea change," *Nature* **439**, 256–260 (2006).
- ⁴⁴T. F. Stocker and D. G. Wright, "Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes," *Nature* **315**, 729–731 (1991).
- ⁴⁵K. Zickfeld, B. Knopf, V. Petoukhov, and H. J. Schellnhuber, "Is the Indian summer monsoon stable against global change?" *Geophys. Res. Lett.* **32**, L15707, <https://doi.org/10.1029/2005GL022771> (2005).
- ⁴⁶Z. Liu and M. Alexander, "Atmospheric bridge, oceanic tunnel, and global climate teleconnections," *Rev. Geophys.* **45**, RG2005, <https://doi.org/10.1029/2005RG000172> (2007).
- ⁴⁷A. Angstrom, "Teleconnections of climate changes in present times," *Geogr. Ann.* **17**, 242–258 (1935).
- ⁴⁸J. Kurths, A. Agarwal, R. Shukla, N. Marwan, M. Rathinasamy, L. Caesar, R. Krishnan, and B. Merz, "Unravelling the spatial diversity of Indian precipitation teleconnections via a non-linear multi-scale approach," *Nonlinear Process. Geophys.* **26**, 251–266 (2019).
- ⁴⁹T. Lenton and J. Ciscar, "Integrating tipping points into climate impact assessments," *Clim. Change* **117**, 585–597 (2013).
- ⁵⁰D. Lemoine and C. P. Traeger, "Economics of tipping the climate dominoes," *Nat. Clim. Change* **6**, 514–519 (2016).
- ⁵¹A. Hastings and D. B. Wysham, "Regime shifts in ecological systems can occur with no warning," *Ecol. Lett.* **13**, 464–472 (2010).
- ⁵²C. Folke, S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, L. Gunderson, and C. Holling, "Regime shifts, resilience, and biodiversity in ecosystem management," *Annu. Rev. Ecol. Evol. Syst.* **35**, 557–581 (2004).
- ⁵³T. Amemiya, T. Enomoto, A. G. Rossberg, T. Yamamoto, Y. Inamori, and K. Itoh, "Stability and dynamical behavior in a lake-model and implications for regime shifts in real lakes," *Ecol. Model.* **206**, 54–62 (2007).
- ⁵⁴M. Scheffer, S. H. Hosper, M.-L. Meijer, B. Moss, and E. Jeppesen, "Alternative equilibria in shallow lakes," *Trends Ecol. Evol.* **8**, 275–279 (1993).
- ⁵⁵B. deYoung, M. Barange, G. Beaugrand, R. Harris, R. I. Perry, M. Scheffer, and F. Werner, "Regime shifts in marine ecosystems: Detection, prediction and management," *Trends Ecol. Evol.* **23**, 402–409 (2008).
- ⁵⁶P. J. Mumby, A. Hastings, and H. J. Edwards, "Thresholds and the resilience of Caribbean coral reefs," *Nature* **450**, 98–101 (2007).
- ⁵⁷V. Brokvík, M. Claussen, V. Petoukhov, and A. Ganopolski, "On the stability of the atmosphere-vegetation system in the Sahara/Sahel region," *J. Geophys. Res.* **103**, 31613–31624, <https://doi.org/10.1029/1998JD200006> (1998).
- ⁵⁸Y. R. Zelnik, S. Kinast, H. Yizhaq, G. Bel, and E. Meron, "Regime shifts in models of dryland vegetation," *Philos. Trans. R. Soc. A* **371**, 20120358 (2013).
- ⁵⁹S. Ellner and P. Turchin, "Chaos in a noisy world: New methods and evidence from time-series analysis," *Am. Nat.* **45**, 343–347 (1995).
- ⁶⁰J. M. Drake and B. D. Griffen, "Early warning signals of extinction in deteriorating environment," *Nature* **467**, 456–459 (2010).
- ⁶¹M. Krkošek and J. M. Drake, "On signals of phase transitions in salmon population dynamics," *Proc. R. Soc. B: Biol. Sci.* **281**, 20133221 (2014).
- ⁶²I. V. Telesh, H. Schubert, K. D. Joehnk, R. Heerkloss, R. Schumann, M. Feike, A. Schoor, and S. O. Skarlatos, "Chaos theory discloses triggers and drivers of plankton dynamics in stable environment," *Sci. Rep.* **9**, 20351 (2019).
- ⁶³C. Trefois, P. M. Antony, J. Gonçalves, A. Skupin, and R. Balling, "Critical transitions in chronic disease: Transferring concepts from ecology to systems medicine," *Curr. Opin. Biotechnol.* **34**, 48–55 (2015).
- ⁶⁴B. Litt, R. Esteller, J. Echaz, M. D'Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, M. Dichter, and G. Vachtsevanos, "Epileptic seizures may begin hours in advance of clinical onset: A report of five patients," *Neuron* **30**, 51–64 (2001).
- ⁶⁵E. Gopalakrishnan, Y. Sharma, T. John *et al.*, "Early warning signals for critical transitions in a thermoacoustic system," *Sci. Rep.* **6**, 35310 (2016).

- ⁶⁶I. Dobson, “Observations on the geometry of saddle node bifurcation and voltage collapse in electrical power systems,” *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **39**, 240–243 (1992).
- ⁶⁷C. Kuehn, E. A. Martens, and D. M. Romero, “Critical transitions in social network activity,” *J. Complex Netw.* **2**, 141–152 (2014).
- ⁶⁸M. Lukosevicius and H. Jaeger, “Reservoir computing approaches to recurrent neural network training,” *Comput. Sci. Rev.* **3**, 127–149 (2009).
- ⁶⁹H. Jaeger, “The “echo state” approach to analyzing and training recurrent neural networks—with an erratum note,” GMD Technical Report No. 148, German National Research Center for Information Technology, Bonn, Germany, 2001.
- ⁷⁰W. Maass, T. Natschläger, and H. Markram, “Real-time computing without stable states: A new framework for neural computation based on perturbations,” *Neural Comput.* **14**, 2531–2560 (2002).
- ⁷¹I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
- ⁷²E. Lorenz, “Deterministic nonperiodic flow,” *J. Atmos. Sci.* **20**, 130 (1963).
- ⁷³Y. Kuramoto, “Diffusion-induced chaos in reaction systems,” *Prog. Theor. Phys. Suppl.* **64**, 346–367 (1978).
- ⁷⁴G. I. Sivashinsky, “On flame propagation under conditions of stoichiometry,” *SIAM J. Appl. Math.* **39**, 67–82 (1980).
- ⁷⁵G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, “Recent advances in physical reservoir computing: A review,” *Neural Netw.* **115**, 100–123 (2019).
- ⁷⁶B. Penkovsky, L. Larger, and D. Brunner, “Efficient design of hardware-enabled reservoir computing in FPGAs,” *J. Appl. Phys.* **124**, 162101 (2018).
- ⁷⁷Y. K. Chembo, “Machine learning based on reservoir computing with time-delayed optoelectronic and photonic systems,” *Chaos* **30**, 013111 (2020).
- ⁷⁸C. Fernando and S. Sojakra, “Pattern recognition in a bucket,” in *Advances in Artificial Life*, edited by W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, and J. T. Kim (Springer, Berlin, 2003), pp. 588–597.
- ⁷⁹J. C. Coulombe, M. C. York, and J. Sylvestre, “Computing with networks of nonlinear mechanical oscillators,” *PLoS One* **12**, 0178663 (2017).
- ⁸⁰K. Harkhoe and G. Van der Sande, “Delay-based reservoir computing using multimode semiconductor lasers: Exploiting the rich carrier dynamics,” *IEEE J. Sel. Top. Quantum Electron.* **25**, 1502909 (2019).
- ⁸¹E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, 2002).
- ⁸²C. Grebogi, E. Ott, and J. Yorke, “Chaotic attractors in crisis,” *Phys. Rev. Lett.* **48**, 1507 (1982).
- ⁸³K. Erguler and M. Stumpf, “Statistical interpretations of the interplay between noise and chaos in the stochastic logistic map,” *Math. Biosci.* **216**, 90–99 (2008).
- ⁸⁴B. Shraiman, C. Wayne, and P. Martin, “Scaling theory for noisy period-doubling transitions to chaos,” *Phys. Rev. Lett.* **46**, 935 (1981).
- ⁸⁵F. Ledrappier and L. S. Young, “Dimension formula for random transformations,” *Commun. Math. Phys.* **117**, 529–548 (1988).
- ⁸⁶L. Yu, E. Ott, and Q. Chen, “Transition to chaos for random dynamical systems,” *Phys. Rev. Lett.* **65**, 2935–2938 (1990).
- ⁸⁷T. Tél, T. Bódai, G. Drótó, T. Haszpra, M. Herein, B. Kaszás, M. Vinczeet *et al.*, “The theory of parallel climate realizations,” *J. Stat. Phys.* **179**, 1496–1530 (2020).
- ⁸⁸S. S. Vallender, “Calculation of the Wasserstein distance between probability distributions on the line,” *Theory Probab. Appl.* **18**, 784–786 (1973).
- ⁸⁹S. Kullback and R. A. Leibler, “On information and sufficiency,” *Ann. Math. Stat.* **22**, 79–86 (1951).
- ⁹⁰S. Chandra, B. R. Hunt, R. Roy, and E. Ott, “Using machine learning to uncover the deterministic dynamics of systems subject to dynamical noise” (to be published).
- ⁹¹F. Christiansen, P. Cvitanovic, and V. Putkaradze, “Spatiotemporal chaos in terms of unstable recurrent patterns,” *Nonlinearity* **10**, 55–70 (1997).