





**Predicting nonsmooth chaotic dynamics by reservoir computing**Lufa Shi, Hengtong Wang , Shengjun Wang , Ruhai Du , and Shi-Xian Qu <sup>\*</sup>*School of Physics and Information Technology, Shaanxi Normal University, Xi'an 710119, China*

(Received 6 April 2023; revised 25 October 2023; accepted 7 December 2023; published 12 January 2024)

Reservoir computing (RC) has been widely applied to predict the chaotic dynamics in many systems. Yet much broader areas related to nonsmooth dynamics have seldom been touched by the RC community which have great theoretical and practical importance. The generalization of RC to this kind of system is reported in this paper. The numerical work shows that the conventional RC with a hyperbolic tangent activation function is not able to predict the dynamics of nonsmooth systems very well, especially when reconstructing attractors (long-term prediction). A nonsmooth activation function with a piecewise nature is proposed. A kind of physics-informed RC scheme is established based on this activation function. The feasibility of this scheme has been proven by its successful application to the predictions of the short- and long-term (reconstructing chaotic attractor) dynamics of four nonsmooth systems with different complexity, including the tent map, piecewise linear map with a gap, both noninvertible and discontinuous compound circle maps, and Lozi map. The results show that RC with the new activation function is efficient and easy to run. It can make perfectly both short- and long-term predictions. The precision of reconstructing attractors depends on their complexity. This work reveals that, to make efficient predictions, the activation function of an RC approach should match the smooth or nonsmooth nature of the dynamical systems.

DOI: [10.1103/PhysRevE.109.014214](https://doi.org/10.1103/PhysRevE.109.014214)**I. INTRODUCTION**

Recently there has been considerable interest in predicting complex nonlinear dynamics with machine learning approaches [1–8]. In particular, reservoir computing (RC) [9,10] has become a paradigm due to its simplicity and high performance. The basic computational substrate of RC is a recurrent neural network called a reservoir, which is actually a high-dimensional nonlinear dynamical system. The reservoir transforms the input data into high-dimensional spatiotemporal patterns, from which the output signal is read out. The main characteristic of RC is that the input weights and the weights of the recurrent connection within the reservoir are not trained, whereas only the readout weights are trained through a simple linear regression. RC has been successfully applied to a wide variety of tasks, such as predicting chaotic time series [2,11–13], producing bifurcation diagrams [8,14], reconstructing chaotic attractors [4,15], forecasting synchronization [16–18], inferring unmeasured state variables [19,20], and so on.

To the best of our knowledge, however, most of the works related to RC focus only on smooth dynamical systems, and nearly no attention has been paid to the nonsmooth dynamical systems. In fact, nonsmooth dynamical systems are ubiquitous in the real world, which usually exhibit abrupt changes of dynamics after gradual and slow integration or evolution. They are a class of very important systems and appear in a wide range of physical, electrical, mechanical, biological, and even economical models, such as relaxation and impact oscillators,

dry frictions, DC converters and switch circuits, neuronal bursting, integrate-and-fire models, and many others. Nonsmooth systems are different from those which are everywhere smooth and differentiable. Their phase spaces are usually divided into individual zones of different dynamical features, and the interfaces between adjacent zones are nondifferentiable. When the dynamical trajectories pass these interfaces, the dynamics changes from one to another and the nonsmooth bifurcation happens [21,22]. These new bifurcations are generally addressed as border-collision bifurcations [23–25]. As a result, especially when both discontinuity and noninvertibility appear [26], a rich class of dynamical phenomena with new characteristics are observed, for example, the coexistence of attractors [27,28], multiple devil's staircases [29], intermittency and crises with new features [30–32], and so on [33–35]. Usually the dynamics and the structure of chaotic attractors in nonsmooth systems are much more complex than those in smooth systems. Therefore, predicting nonsmooth dynamical systems is a great challenge. Our numerical work shows that the conventional RC with a hyperbolic tangent activation function can hardly model and forecast the dynamics of the simplest piecewise linear system like the tent map. New schemes are expected to solve this problem.

In the process of the machine learning community's work to improve the efficiency and performance of RC, some work has demonstrated that neural networks should respect the property of learned systems. For example, it has been reported that matching the RC symmetry with the symmetry of the data being processed can considerably improve the RC performance [36]. The deep neural network that inherently respects the scaling and symmetry properties of the original governing Navier-Stokes equations has been developed to simulate the

<sup>\*</sup>Corresponding author: [sxqu@snnu.edu.cn](mailto:sxqu@snnu.edu.cn)

fluid dynamics more efficiently [37]. In addition, Hamiltonian neural networks have also been tailored by embedding the concepts of energy conservation and volume preservation arising from an underlying Hamiltonian function into neural networks [38,39]. Such neural networks are usually known as physics-informed neural networks [40,41], which are emerging as very promising machine learning approaches due to better accuracy, faster training, and improved generalization. These works show that a well-adapted neural network should be constrained to respect the symmetry, invariance, or conservation principles originating from the physical laws that govern the observed data.

Actually, still other kinds of dynamical properties should be considered, especially when one tries to coin a RC scheme to predict the dynamics of nonsmooth systems. They are the nonlinearity and the smoothness of the dynamics. It may be a reasonable requirement for the RC internal dynamics to match at least the nonsmoothness of the dynamics hidden in the input. Thus, for the nonsmooth systems involved in the current work, our point of penetration is the activation function since it governs the RC internal dynamics. We are thinking of employing nonsmooth activation functions to replace the smooth ones. The good results in forecasting the dynamics and reconstructing chaotic attractors are obtained by applying the RC with the nonsmooth activation function to the tent map, the piecewise linear map with a gap [28,32], the compound circle maps describing an electric relaxation oscillation [26], and the two-dimensional nonsmooth Lozi map.

The paper is organized as follows. In Sec. II the architecture and details of the RC are described. In Sec. III an activation function with piecewise linear nature is proposed by analyzing the failure of RC with hyperbolic tangent functions. The results and discussion appear in Sec. IV, and Sec. V is the conclusion.

## II. THE RESERVOIR MODEL

### A. Reservoir computing

In RC framework, the reservoir is actually a sparse neuron network internally connected with linking matrix  $\mathbf{A} \in \mathbb{R}^{D_r \times D_r}$  [1,2,4]. The nonzero entries of  $\mathbf{A}$  are drawn from uniformly distributed random numbers in  $[0,1]$ , and then scaled so that the absolute value of the largest eigenvalue of the matrix, the spectral radius, equals  $\rho$ . In this work the sparsity of  $\mathbf{A}$  is adjusted so that the network has an average degree of six [4]. The internal dynamical state of the network at time step  $n$  is stored in vector  $\mathbf{r}(n) \in \mathbb{R}^{D_r \times 1}$ . Here a proper value of  $D_r$  is set to match the dynamical complexity. In conventional RC, an input vector  $\mathbf{u}(n)$  of dimension  $D$  is coupled to the  $D_r$ -dimensional state space of the reservoir through weight matrix  $\mathbf{W}_{in} \in \mathbb{R}^{D_r \times D_{in}}$ , whose elements are chosen from the random numbers distributed uniformly in the interval  $[-\beta, \beta]$ . The evolution of the internal dynamics of the reservoir is governed by

$$\mathbf{r}(n+1) = \mathbf{G}[\mathbf{A}\mathbf{r}(n) + \mathbf{W}_{in}\mathbf{u}(n)], \quad (1)$$

where  $\mathbf{G}$  is the activation function. The  $D$ -dimensional output vector  $\mathbf{v}(n+1)$  can be obtained by

$$\mathbf{v}(n+1) = \mathbf{W}_{out}[\mathbf{r}(n+1), \mathbf{P}], \quad (2)$$

where  $\mathbf{W}_{out} \in \mathbb{R}^{D_{out}}$  maps the  $D_r$ -dimensional dynamical states of the reservoir into  $D_{out}$ -dimensional output states, and  $\mathbf{P}$  is the output matrix [2], which contains  $D_{out} \times D_r$  adjustable parameters (or weights) obtained by linearly fitting Eq. (2) into the target  $\mathbf{v}_d(n+1)$  during the training process. In this paper we assume that  $\mathbf{W}_{out}[\mathbf{r}, \mathbf{P}]$  depends linearly on  $\mathbf{P}$ :  $\mathbf{W}_{out}[\mathbf{r}, \mathbf{P}] = \mathbf{P}\mathbf{r}$ . For simplicity we set  $D_{in} = D_{out} = D$ .

In performing RC to predict the dynamics, the input time series for training the system are actually the dynamical trajectories  $\{\mathbf{u}(t), t = 0, \dots, N_W + N_T\}$ , where  $\mathbf{u}(0)$  is the initial state fed into RC, before which 20000 transient states are eliminated so that it is on chaotic attractors. The dynamical trajectories are produced by

$$\mathbf{u}(t) = \mathbf{F}(\mathbf{u}(t-1)), \quad \text{for } t > 0, \quad (3)$$

where  $\mathbf{F}(\cdot)$  is formally a time-discrete mapping, i.e., a real map for a time-discrete system or iteration solutions of time-continuous systems. The time series is fed into the reservoir to train the machine, and accordingly a time series  $\{\mathbf{r}(t), t = 1, \dots, N_W + N_T\}$  of the internal states is produced. The first  $N_W$  steps are the so-called warm-up phase, which are usually dropped to prevent the influence of arbitrary initial states of the reservoir. Thus we set  $n = t - N_W$ , and  $\mathbf{u}(0) = \mathbf{u}(t = N_W)$ , and then we have input states  $\{\mathbf{u}(n), n = 0, \dots, N_T - 1\}$  and internal states  $\{\mathbf{r}(n), n = 1, \dots, N_T\}$ . The corresponding target is  $\{\mathbf{v}_d(n) = \mathbf{u}(n), n = 1, \dots, N_T\}$ . The training is finished by determining the output matrix  $\mathbf{P}$  via ridge regression, which minimizes the following quantity:

$$\sum_{n=1}^{N_T} \|\mathbf{W}_{out}[\mathbf{r}(n), \mathbf{P}] - \mathbf{v}_d(n)\|^2 + \alpha \|\mathbf{P}\|^2, \quad (4)$$

where  $\alpha$  is a small positive regularization constant used to avoid overfitting, and  $\|\cdot\|$  is the modulus of a matrix. Once the mapping  $\mathbf{W}_{out}$  is obtained, the RC switches into the autonomous prediction phase.

In the forecasting phase, we replace  $n$  by  $n - N_T$  and get the starting state  $\mathbf{v}(0) = \mathbf{u}(N_T)$ . When we feed this  $\mathbf{v}(0)$  into the RC, then the corresponding prediction  $\mathbf{v}(1)$  is obtained by Eq. (2). Feeding this new prediction back to the RC, we get the next prediction  $\mathbf{v}(2)$ . Repeating this process sequentially, the RC can conduct one-step-ahead predictions autonomously.

### B. Criteria for the efficiency of prediction

To explore the capability of the reservoir in forecasting the chaotic dynamics, we inspect two kinds of predictions. One is the short-term prediction, and the other is the long-term prediction. Suppose the predicted time series for an arbitrary time  $N$  are  $\{\mathbf{v}(n), n = 1, \dots, N\}$ , and the corresponding true counterparts produced by the map in Eq. (3) with the same starting state are  $\{\mathbf{v}_d(n), n = 1, \dots, N\}$ .

For a short-term prediction, the measure of forecasting quality is the valid prediction time (VPT) denoted by  $T_\lambda$ , which is the longest duration of time, i.e.,  $\max(N)$ , after which the normalized error  $E(N)$  of prediction exceeds a threshold  $\varepsilon$ . It is defined by

$$T_\lambda \equiv \frac{\max(N)}{1/\lambda_m}, \quad \text{if } E(N) \leq \varepsilon, \quad (5)$$

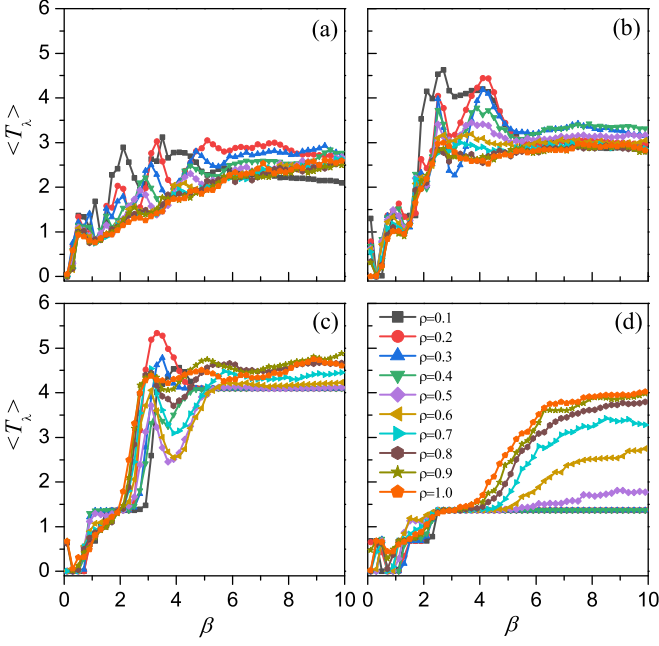


FIG. 1. Hyperparameter dependence of the mean valid prediction times for the tent map by RC when  $b = 1$ . (a)  $\alpha = 10^{-8}$ ; (b)  $\alpha = 10^{-6}$ ; (c)  $\alpha = 10^{-4}$ ; (d)  $\alpha = 10^{-2}$ .

where  $\lambda_m$  is the maximum Lyapunov exponent, and thus  $T_\lambda$  is in the unit of Lyapunov time  $1/\lambda_m$ . We set  $\varepsilon = 0.1$ . The normalized error is defined by

$$E(N) \equiv \frac{\|\mathbf{v}(N) - \mathbf{v}_d(N)\|}{\sqrt{\langle \mathbf{v}_d^2 \rangle}}, \quad (6)$$

where the mean square of the targets is given by

$$\langle \mathbf{v}_d^2 \rangle = \frac{1}{N} \sum_{n=1}^N \|\mathbf{v}_d(n)\|^2. \quad (7)$$

A long-term forecasting is obtained if one prolongs the previous short-term prediction until a large  $N_P$  steps. The predicted dynamical trajectories are  $\{\mathbf{v}(n), n=1, \dots, N_P \gg 1\}$ . Based on those trajectories, one can reconstruct the chaotic

attractor in the phase space. To intuitively display how far the predicted dynamical trajectories are from the actual ones in the long-term prediction, we employ the return map analysis since most of the systems involved in this work are one-dimensional maps. The actual return mapping function is drawn by blue lines for one-dimensional systems. Accordingly, in the same plot, we mark pairs of successively predicted trajectories  $\{(\mathbf{v}(n-1), \mathbf{v}(n)), n=1, \dots, N_P\}$  by red circles, as shown in Figs. 2, 5, 7, 9. Similarly, the true trajectories and the predicted ones are drawn by blue and red dots, respectively, as shown in Fig. 10. We can visually inspect whether those predicted iterative trajectories fall on the true maps in the phase space domain occupied by the chaotic attractor. Quantitatively, the measure for the difference between the predicted mapping points  $\mathbf{v}(n)$ s and the corresponding theoretical map points  $\mathbf{F}(\mathbf{v}(n-1))$ s for  $n=1$  to  $N_P$  is introduced, which is defined by the root mean square error (RMSE)

$$\sigma = \left\{ \frac{1}{N_P - 1} \sum_{n=1}^{N_P} [\mathbf{v}(n) - \mathbf{F}(\mathbf{v}(n-1))]^2 \right\}^{1/2}. \quad (8)$$

It reveals globally how far the prediction is from the real one in the long-term forecasting, or the precision in reconstructing the chaotic attractor.

We have to admit that the return map analysis of the predicted trajectories ignores some dynamical details of the chaotic orbit, even though it provides us with an intuitive way to measure the forecasting quality of RC. To compensate for this shortcoming, the largest Lyapunov exponent  $\lambda_R$  of the autonomous RC (or the projected attractor) is used to characterize the dynamical feature and evaluate whether RC successfully replicates the long-term evolution or the chaotic attractors of dynamical systems. It is calculated by the algorithm in Ref. [4].

### C. Hyperparameter optimization

The hyperparameter optimization of the RC is performed by the coarse grid sweep in the parameter regimes of  $\rho$ ,  $\beta$ , and  $\alpha$ . Here the increment of  $\rho$  and  $\beta$  is 0.1, and  $\alpha$  are chosen as  $\alpha = 10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}$ , and  $10^{-2}$ . As is well known,

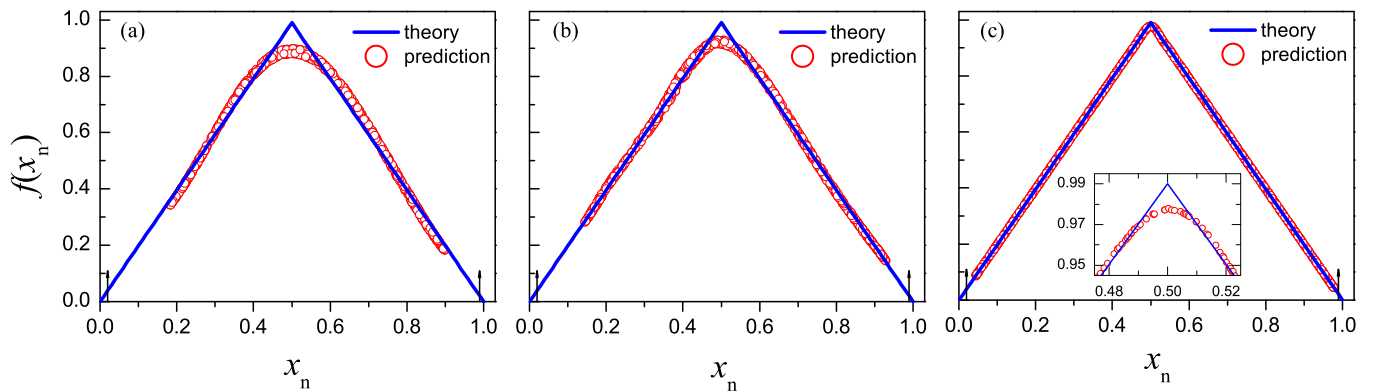


FIG. 2. Reconstructed attractors of the tent map by RC with  $\tanh(bx)$ . The lines represent the true map, and circles are the predictions. The pairs of arrows confine the phase-space regions occupied by attractors. (a)  $b = 1$ ,  $\sigma = 0.03$ , and  $T_\lambda = 6.15$ ; (b)  $b = 10$ ,  $\sigma = 0.015$ , and  $T_\lambda = 6.83$ ; (c)  $b = 100$ ,  $\sigma = 0.00095$ , and  $T_\lambda = 7.51$ . The insert is the enlarged plot of the map near the peak.

different random realizations of  $\mathbf{A}$  and  $\mathbf{W}_{\text{in}}$  for a set of given  $\rho$  and  $\beta$  may seriously affect the prediction of RC, thus the characteristic quantities are computed using the average over 100 runs with different random realizations of them.

We have to mention that the size of the reservoir  $D_r$ , the warm-up time  $N_W$ , the training time  $N_T$ , and the long-term prediction time  $N_P$  are also adjustable parameters, which are often empirically determined. In general, a more complex attractor structure requires a larger network size ( $D_r$ ), and more nodes ( $D_r$ ) may result in higher prediction accuracy. Meanwhile, larger size means more computation time. Therefore, the choice of  $D_r$  is usually the result of the tradeoff between accuracy and computation time. In the literature related to RC, the size of the reservoir takes a small value for simple systems (such as  $D_r = 300$  or  $500$ ), and  $N_W = 50$  or  $100$ . The basic requirement for the training time  $N_T$  is  $N_T > D_r$  to avoid underfitting. The value of the long-term prediction time  $N_P$  should be large enough to make the trajectories visit most of the phase space domain occupied by the attractor.

### III. THE RESERVOIR COMPUTER FOR NONSMOOTH DYNAMICS

#### A. RC failure with smooth activation function

One of the simplest nonsmooth dynamical systems that displays chaos might be the tent map, which is nonsmooth but continuous, defined by

$$x_{n+1} = \begin{cases} rx_n, & 0 \leq x_n \leq \frac{1}{2} \\ r(1 - x_n), & \frac{1}{2} \leq x_n \leq 1 \end{cases}. \quad (9)$$

Here  $r$  is the control parameter. It is topologically conjugated to the logistic map and thus shares many common dynamical features. The remarkable difference is the nondifferentiability at the peak. The Lyapunov exponent  $\lambda$  of the system is defined through

$$\lambda \equiv \lim_{N \rightarrow \infty} \left\{ \frac{1}{N} \sum_{n=0}^{N-1} \ln |f'(x_n)| \right\}, \quad (10)$$

where  $f'(\cdot)$  is the derivative of the mapping function. Clearly,  $|f'(x_n)| = r$  for all  $x_n$ . Hence, the analytical expression for the Lyapunov exponent reads  $\lambda = \ln r$ . In this work we set  $r = 1.98$ ,  $\lambda = 0.683$ , and the system is in the chaotic state.

In the conventional RC to predict the chaotic dynamics of smooth systems, the hyperbolic tangent function is often employed as an activation function [2,3,5–7]. In Ref. [42] the argument of the hyperbolic tangent function is rescaled by a factor  $b$ ,

$$G(x) \equiv \tanh(bx). \quad (11)$$

The factor  $b$  influences the nonlinearity of the function and thus may affect the quality of prediction. Rescaling the argument is equivalent to adjusting hyperparameters  $\rho$  and  $\beta$  simultaneously. Now we simply apply the RC with this activation function to predict the dynamics and reconstruct the chaotic attractor of the tent map in Eq. (9). Three different values of  $b$  are chosen:  $b = 1$ ,  $10$ , and  $100$ . The parameters of the reservoir are  $D_r = 300$ ,  $N_W = 50$ ,  $N_T = 450$ , and  $N_P = 1000$ . The results are described in the following.

The mean valid prediction times when  $b = 1$  at all swept parameters are summarized in Fig. 1. Obviously, the quality of short-term prediction is not too bad where the mean VPT denoted by  $\langle T_\lambda \rangle$  is around five Lyapunov times. Comparing all the plots, the best result appears in the parameter set in Fig. 1(c). The hyperparameters are  $\rho = 0.2$ ,  $\beta = 3.3$ , and  $\alpha = 10^{-4}$ , respectively. For the best random network realization at this set of hyperparameters, the VPT is 6.15 Lyapunov times and the RMSE for reconstructing chaotic attractor in the long-term prediction keeps a large value,  $\sigma = 3.0 \times 10^{-2}$ . The corresponding reconstructed trajectories of the chaotic attractor are drawn by the red circles in Fig. 2(a). One can see that at the middle portion of the left and right halves of the map, the predicted results agree very well with the theory. Inside the chaotic attractor, there are still two pretty large regions near the two arrows in the figure, having not been visited by the predicted trajectory. Near the peak of the map, however, not only do the prediction results severely deviate from the theory, but also the map segment coined by the trajectories of the reconstructed attractor (or reconstructed map) exhibits smooth variation near the peak rather than a nonsmooth tip as in the original map. The reconstructed map shows a smooth peak. The failure of the RC in the long-term prediction can also be verified by the rather large difference between the largest Lyapunov exponent  $\lambda_R = 0.49$  of the autonomous RC and  $\lambda = 0.683$  of the target attractor. The relative error is up to 28%. Thus the hyperbolic tangent function when  $b = 1$  is not a good choice of the activation function for RC when predicting the nonsmooth dynamics, and especially when replicating chaotic attractor.

When the scaling factor increases to  $b = 10$ , the optimized hyperparameters are  $\rho = 0.1$ ,  $\beta = 1.5$ , and  $\alpha = 10^{-4}$ , respectively. The longest VPT for a special network realization is 6.83 Lyapunov times. The long-term trajectories are the circles plotted in Fig. 2(b) to replicate the map inside the chaotic attractor, where the RMSE reduces to  $\sigma = 1.5 \times 10^{-2}$ . The similar results are obtained. The improvement on the results is observed, e.g., the deviation of predicted trajectories near the peak from the theoretical counterparts is decreased, and the curvature becomes larger than that in Fig. 2(a). In addition, the improvement is reflected by the fact that the largest Lyapunov exponent of the reservoir system in the prediction phase is  $\lambda_R = 0.53$ , and the relative error with respect to that of the system is reduced to 22%.

As the scaling factor goes to  $b = 100$ , the optimized hyperparameters are  $\rho = 0.3$ ,  $\beta = 0.3$ , and  $\alpha = 10^{-10}$ , respectively. The longest VPT for a special network realization is 7.51 Lyapunov times. The long-term trajectories are represented by red circles plotted in Fig. 2(c) to replicate the map inside the attractor, and the RMSE drastically is reduced to  $\sigma = 9.5 \times 10^{-4}$ . The trajectories visit nearly the whole region of the attractor and fall nicely on the map function. The deviation from the theoretical counterparts becomes much smaller than those in Figs. 2(a) and 2(b). The predicted map near the peak becomes much sharper than in the previous two cases, so that one cannot even witness the smoothness of it without carefully inspecting the inserted plot in Fig. 2(c). Another piece of evidence for the better prediction corresponding to Fig. 2(c) is that the largest Lyapunov exponent  $\lambda_R = 0.66$ , which is pretty close to that of the target attractor, where the



relative error significantly is reduced to 3.4%. Yet the RC cannot replicate the nonsmooth peak even though the prediction in other regions of the attractor becomes much better.

### B. Nonsmooth activation function and modified RC

The above results suggest that RC with smooth activation functions performs well in the short-term prediction of chaotic dynamics in the tent map. However, in the long-term prediction, it cannot capture the nonsmooth nature of the map even though the improvement on the precision is more significant than the short-term counterpart when the scaling factor  $b$  increases. The smoothness of the reconstructed map decreases as  $b$  increases. How can we understand these conflicted results? The pretty good performance of the short-time prediction can be explained by the fact that one cannot observe the difference between the nonsmooth from smooth dynamics when the trajectories are far from the nonsmooth border or the control parameter does not approach its critical value of the border-collision bifurcation at which the dynamical behavior undergoes an abrupt transition. The data used to train the reservoir computer are produced by the tent map, which is a piecewise linear map and is nondifferentiable at the peak. What is the origin of the smoothness and nonlinearity in the reconstructed map near the peak?

To answer the question, we go back to the working principle of RC, where the time series fed into the reservoir are transformed to spatiotemporal patterns in the high-dimensional state space, i.e., making a high-dimensional duplication of the input data. It has been proven that the artificial neural network is a universal function approximator [43,44]. However, this does not necessarily mean that the reservoirs with all activation functions will perform equally well. That is to say, the degree of approximation depends on the property of the activation function. In the RC here, both the input and output functions are linear, thus the activation function is the only source of nonlinearity in the system, which might contribute to the smoothness of the reconstructed attractor near the peak.

Moreover, it is necessary to inspect the hyperbolic tangent functions in Fig. 3. The function tends to  $\pm 1$  asymptotically as  $x$  goes to  $\pm\infty$ , and the slope at  $x=0$  is  $b[1 - \tanh^2(bx)]|_{x=0} = b$ . As an example, we consider the case of  $b=1$  (the solid black line in the figure) first. The domain of the argument can be roughly divided into five pieces under certain precision, i.e., the linear zone with slope  $b$ , two saturation zones, and two curved zones between the linear and saturation zones. The nonlinearity and smoothness of the activation function mainly are attributed to the two curved zones that bend outward. As  $b$  increases, their curvatures increase and the curved zones become sharper, which is very similar to the behavior of the reconstructed map near the peak. Hence, the two curved regions may most probably be the origin of the nonlinearity and smoothness of the duplicated map near the peak. Considering the nonsmooth nature of the tent map, however, the nonsmoothness should be introduced into the activation function. To do so, let the tangent of  $\tanh(bx)$  intersects with two asymptotical saturation lines at the locations marked by two arrows in Fig. 3, and then link it with the saturate lines at the intersections, then we immediately

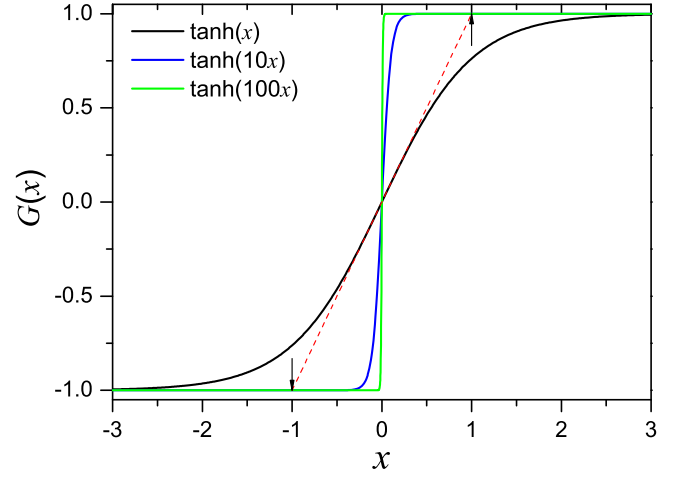


FIG. 3. Activation functions with different scaling factors. The dashed line is the tangent of  $\tanh(x)$ .

get a piecewise linear activation function, which is generally defined by

$$G(x) \equiv \begin{cases} -1, & x < -1/b \\ bx, & |x| \leq 1/b \\ 1, & x > 1/b \end{cases} \quad (12)$$

Although it has the piecewise linear nature, this function satisfies the requirement on the continuous sigmoidal function and thus can be used to approximate any function in echo-state network realizations [43,44]. Therefore, we choose it as the activation function and build a modified RC in the current work. Without loss of generality, we set  $b=1$ . We have to emphasize that the new activation function does not change the echo-state property or stability of the RC significantly, and the detailed analysis is in the Appendix. There already has been a simpler nonsmooth activation function, which is the so-called rectified linear activation function,  $\text{ReLU}(x) = \max(0, x)$ . It is usually designed to solve the problem of vanishing gradient in deep learning. We established a RC with this ReLU to predict the tent map but often got trajectories beyond the domain of the map function, and thus cannot replicate the map. The reason may be that there is no upper bound for this function. Then we turned to the above mentioned procedure to look for one as a substitute. We want to emphasize that the nonsmoothness of RC with this activation function defined in Eq. (12) may match the same nature of the nonsmooth dynamical systems. This can be addressed as a kind of physics-informed neural network. The results of its applications are described in the next section.

## IV. PREDICTING CHAOTIC DYNAMICS OF NONSMOOTH SYSTEMS WITH THE MODIFIED RC

### A. Tent map

The first application is to the tent map with the nonsmooth but continuous natures, which has the simplest attractor structure. Here the parameters of the reservoir are  $D_r = 300$ ,  $N_w = 50$ ,  $N_T = 450$ , and  $N_P = 1000$ . To compare with those in Fig. 1, drawn in Fig. 4 are the mean valid prediction times obtained by the same sweep procedure described in

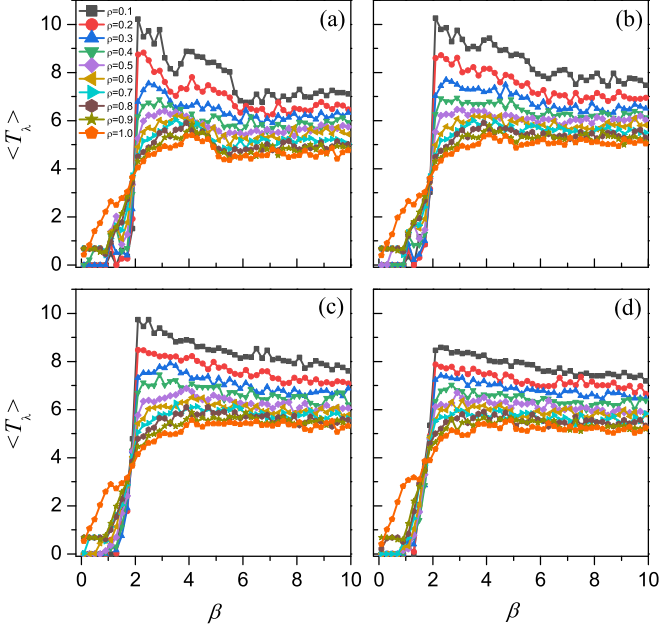


FIG. 4. Hyperparameter dependence of the mean valid prediction times for the tent map by RC with piecewise linear activation function. (a)  $\alpha = 10^{-10}$ ; (b)  $\alpha = 10^{-8}$ ; (c)  $\alpha = 10^{-6}$ ; (d)  $\alpha = 10^{-4}$ .

the last section. The significant increase in the mean VPTs is observed, where the longest VPT is beyond 10 Lyapunov time, and even the worst is larger than that in Fig. 1. One may also find that the VPTs become sensitive to parameter  $\rho$  and behave in a bit of regular dependence on  $\rho$ , which may stem from the piecewise-linear feature of the activation function. Furthermore, the VPTs are very close to 0 as  $\rho$  and  $\beta$  are small, because the distribution of the argument of the activation function produced in the iteration of the reservoir is confined in a very narrow region, and the RC cannot grasp the dynamical nature of the tent map. As  $\beta$  increases to the vicinity of 1.9, the VPTs increase significantly, where the distribution is extended to much broader regions, even covering the nonsmooth intersections of the activation function, which is essential for the RC to model nonsmooth dynamical systems.

The optimized hyperparameters by comparing the results shown in Fig. 4 are  $\rho = 0.1$ ,  $\beta = 2.1$ , and  $\alpha = 10^{-10}$ , respectively. For the best random network realization, the VPT is 13.0 Lyapunov times, and the RMSE for reconstructing the map in the long-term prediction is  $\sigma = 1.0 \times 10^{-5}$ , which is smaller than that in the last section by an order of 2. The corresponding reconstructed trajectories of the chaotic attractor are shown by the circles in Fig. 5. One clearly sees that the prediction fits the theory very well, and all the trajectories fall in the map exactly. Very important is that the replicated map near the peak exhibits nonsmoothness, as shown by the insertion. Furthermore, the largest Lyapunov exponent of the RC is  $\lambda_R = 0.68$ , which is very close to the actual value 0.683 of the target attractor, and the relative error is only 0.44%. These facts also confirm the success of the RC in replicating the long-term dynamics. Therefore, we may conclude that the RC with the proposed activation function has extracted the nonsmooth but continuous natures of the map, and then

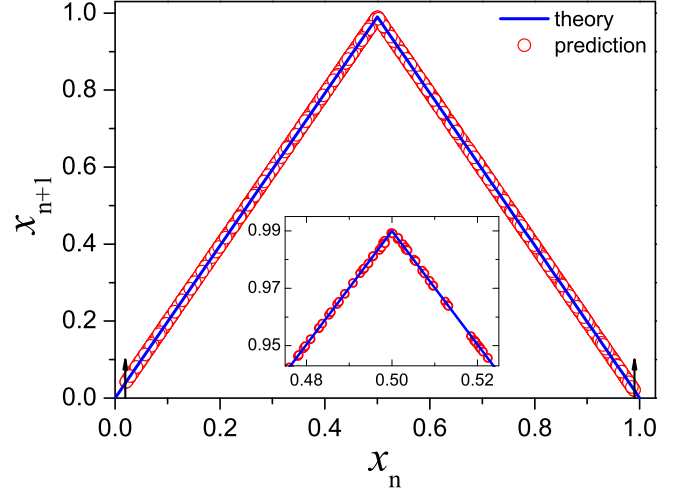


FIG. 5. Reconstruction of the tent map by RC with piecewise linear activation function. The line represents the tent map, and the circles are the predicted counterparts.

the RC can precisely predict its dynamics and perfectly reconstruct the attractor.

On the other hand, one may ask whether RC equipped with the proposed nonsmooth activation function also performs better in predicting smooth dynamics. To answer this question, we employ the logistic map as an illustration of a smooth system and predict its dynamics via RC with nonsmooth and smooth activation functions, respectively. The control parameter of the map is 3.98 where the system is in the chaotic state, and the Lyapunov exponent is  $\lambda = 0.616$ . The nonsmooth activation function is the piecewise linear function proposed in this work, and the smooth counterpart is the conventional hyperbolic tangent function. For the short-term forecasting, the mean VPTs,  $\langle T_\lambda \rangle$ , and best VPTs are calculated. For the long-term forecasting, the RMSEs,  $\sigma$ , and largest Lyapunov exponents are also calculated. All results are listed in Table I for comparison. Obviously, the RC with the conventional hyperbolic tangent activation function performs much better than the one equipped with the nonsmooth activation function in predicting the dynamics of the smooth logistic map. This implies it is important that the smoothness of the activation function should match that of the systems in forecasting their dynamics, which coincides with the spirit of physics-informed neural networks.

### B. Piecewise linear map with a gap

The second nonsmooth dynamical system is a piecewise linear discontinuous map, defined by

$$x_{n+1} = f_i(x_n) = k_i x_n + b_i, \quad i = 1, 2, 3, 4 \quad (13)$$

TABLE I. Comparison of measures for short- and long-term predictions of logistic system by RC with different activation functions.

$G$	$\langle T_\lambda \rangle$	Best $T_\lambda$	$\sigma$	$\lambda_R$	$ 1 - \lambda_R/\lambda $
Nonsmooth	5	6.78	$3 \times 10^{-3}$	0.59	4.2%
Smooth	9	12.3	$1.3 \times 10^{-5}$	0.61	0.97%

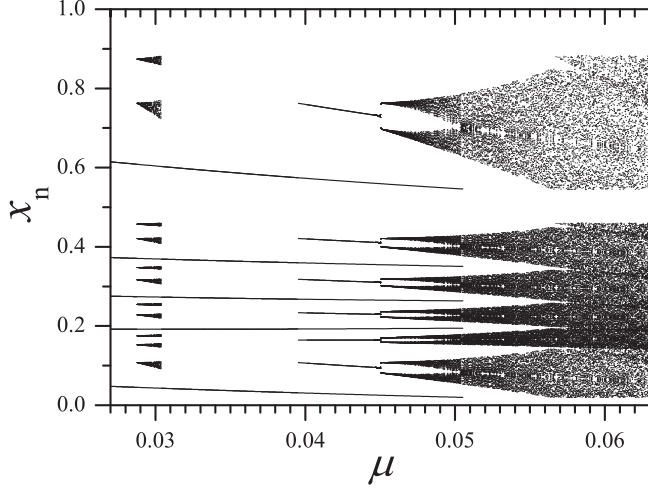


FIG. 6. Bifurcation diagram of the piecewise linear map.

with

$$\begin{aligned} k_1 &= -\frac{\mu}{x_b - x_A}, & b_1 &= y_A, & x &\in [x_A, x_b]; \\ k_2 &= \frac{y_C - b_1 + \mu}{x_g - x_b}, & b_2 &= y_b - k_2 x_b, & x &\in [x_b, x_g]; \\ k_3 &= \text{const}, & b_3 &= \text{const}, & x &\in [x_g, x_F]; \\ k_4 &= \text{const}, & b_4 &= \text{const}, & x &\in [x_F, x_G], \end{aligned}$$

where  $\mu$  is the control parameter. The other parameters are  $y_A = 0.203921$ ,  $y_C = 0.460000$ ,  $y_G = y_A$ ,  $x_A = 0$ ,  $x_b = 0.107663$ ,  $x_g = 0.350000$ ,  $x_F = 0.497121$ ,  $x_G = 1$ ,  $k_3 = 3.07055$ ,  $b_3 = -0.530165$ ,  $k_4 = 0.405507$ , and  $b_4 = -0.201586$ . It has a discontinuity at  $x_g$  and a noninvertible zone in  $[f(x_b), y_A]$ , which is the main difference from the tent map. The interplay between the discontinuity and noninvertibility may result in new kinds of bifurcations, such as the coexistence of periodic attractors [28], and evidence for the hole-induced crisis [32], shown by the bifurcation diagram in Fig. 6. Here we are interested in the band-6 chaos, which is due to the merging of band-12 chaos induced by the discontinuous bifurcation from a period-6 orbit. The attractor structure is more complex than that of the tent map. One can see that there are forbidden regions where the trajectory cannot visit. The borders of the chaotic attractor are determined by the minimum  $(x_b, y_b)$  and its forward iterations, shown by the green dashed lines in Fig. 7.

To forecast the chaotic dynamics by learning the time series produced by the above map when the band-6 chaos appears, we set  $\mu = 0.055$ . The Lyapunov exponent is  $\lambda = 0.085$ . The RC parameters are  $D_t = 300$ ,  $N_W = 50$ ,  $N_T = 950$ , and  $N_P = 1000$ . The hyperparameters are set to  $\rho = 0.5$ ,  $\beta = 2.9$ , and  $\alpha = 10^{-8}$ . The mean VPT can reach almost seven Lyapunov times, and the longest is 11.2 Lyapunov times in the short-term prediction. The reconstructed map within the basin of the chaotic attractor (or the reconstructed attractor) is represented by the red circles in Fig. 7. Obviously, the trajectories fit the actual map very well with the precision  $\sigma = 1.3 \times 10^{-5}$ . Moreover, the boundary of the reconstructed attractor coincides with the actual boundary, and the largest

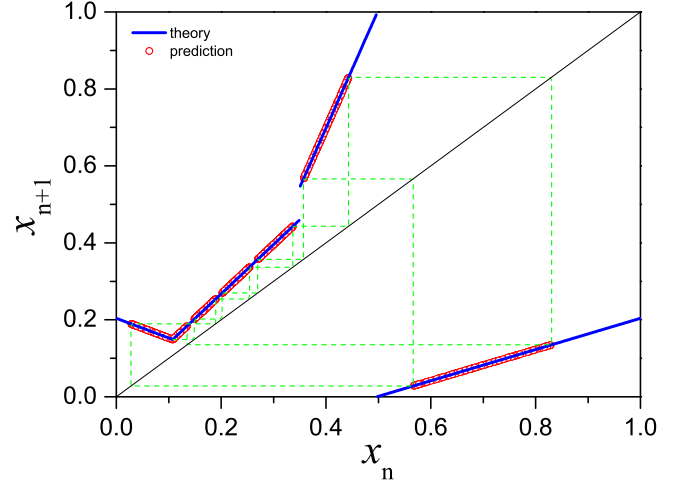


FIG. 7. Reconstructed map (chaotic attractor) of the piecewise linear map with a gap. Boundaries of the chaos bands are marked by the dashed lines that are the forward iterations of the minimum.

Lyapunov exponent  $\lambda_R = 0.082$  of the autonomous reservoir is very close to that of the target attractor, where the relative error is 3.5%. These results imply that the RC has successfully learned the nonsmooth and discontinuous natures of the piecewise linear map with a gap, and made a perfect reconstruction of the band-6 chaotic attractor.

### C. Both noninvertible and discontinuous compound circle maps

Illustrated in this section is the application to the compound circle maps, which describe an electric relaxation oscillation between the sine-modulated up- and the constant down-threshold in a thyatron circuit [26]. It has an implicit iteration function form,

$$[F_2(x_{n+1}) - F_1(x_n)], \quad \text{mod } 1 = 0 \quad (14)$$

with

$$F_i(x) = x + A_i + B_i \ln[C_i + \sin(2\pi x)], \quad i = 1, 2, \quad (15)$$

where  $x_n$  and  $x_{n+1}$  are the phases of modulation signal, at which two adjacent relaxation oscillations reach the up-threshold, respectively. Other quantities are  $A_1 = \omega RC \ln \{v_0/[U_{\min} - (R/R_1)E]\}/2\pi$ ,  $B_1 = \omega RC/2\pi$ ,  $C_1 = [U_{\max} - (R/R_1)E]/U_0$ ,  $A_2 = \omega R_1 C \ln[U_0/(E - U_{\min})]/2\pi$ ,  $B_2 = \omega R_1 C/2\pi$ ,  $C_2 = (E - U_{\max})/U_0$ , and  $R = R_1 R_2/(R_1 + R_2)$ , where  $\omega$  and  $U_0$  are circular frequency and amplitude of the modulation signal, respectively.  $U_0$  is selected as the control parameter, and the others are the parameters of the electronic circuit. Here we set  $\omega = 900$  Hz,  $E = 99.0$  V,  $R_1 = 200$  k $\Omega$ ,  $R_2 = 170$  k $\Omega$ ,  $C = 1.65 \times 10^{-8}$  F,  $U_{\max} = 64$  V, and  $U_{\min} = 51.2$  V.

In this work the map is tuned to its both noninvertible and discontinuous regime, where it is actually a piecewise nonlinear map with nonsmoothness and discontinuity. Plotted in Fig. 8 is the bifurcation diagram. There is a band-3 chaotic attractor near  $U_0 = 12.57842$ , at which the Lyapunov exponent is  $\lambda = 0.056$ . The boundaries of the band-3 chaotic attractor are the forward iterations through the minimum of the map function, denoted by the dot-dashed lines, and the  $(x_g, f(x_g^+))$  (up border of the gap) and its forward iterations,

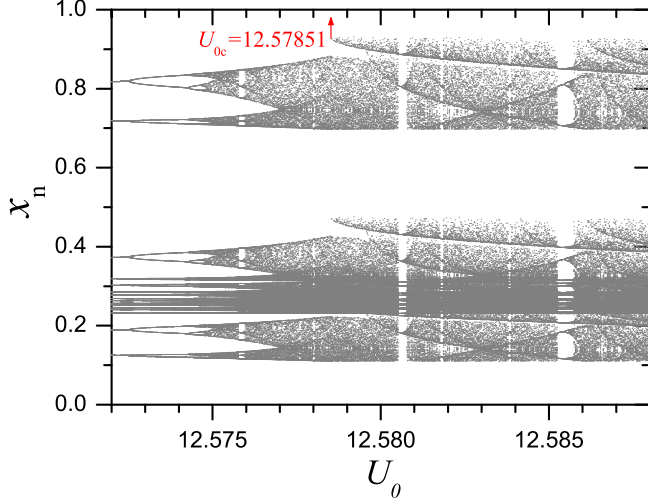


FIG. 8. Bifurcation diagram of the compound circle maps.

denoted by the dashed lines in Fig. 9(a). The attractor stems from a route of a series of periodic doubling cascades via the period-16 orbit, merging of chaotic attractors, and intermittencies. When the control parameter is beyond the critical value,  $U_{0c} = 12.57851$ , the chaotic attractor suddenly expands into five pieces, which is a kind of crisis with striking characteristics. It is induced by the collision between the chaotic attractor and the leaking hole in the high-order map originating from the interaction between discontinuity and noninvertibility, and thus addressed as hole-induced crisis [31,32]. Then a band-5 chaotic attractor is observed at  $U_0 = 12.57950$  after the crisis, and its Lyapunov exponent is  $\lambda = 0.078$ . The borders of the band-5 chaotic attractor are the forward iterations of the minimum of the map function, denoted by the dot-dashed lines, and the  $(x_g, f(x_g^+))$  (up border of the gap) and  $(x_g, f(x_g^-))$  (down border of the gap) and their forward iterations, denoted by the dashed lines and the dotted lines in Fig. 9(b), respectively. In both cases the chaotic attractors are multiple bands divided by the forbidden regions in the phase space, which contributes to the complexity of the attractors. The structures of the attractors are much more complicated than that in the last subsection.

To simulate the dynamics in the band-3 chaotic state, the dimension of the reservoir is set to  $D_r = 1500$ , the number of the time steps for the warmup phase is  $N_w = 100$ , the number of data for training is  $N_T = 1900$ , and the number of the steps in the long-term prediction is  $N_P = 2000$ . The hyperparameters are  $\rho = 0.1$ ,  $\beta = 3.9$ , and  $\alpha = 10^{-8}$ . In the short-term prediction, the mean VPT is 6.05 Lyapunov times and the longest VPT is 9.72 Lyapunov times. After the crisis, we set  $D_r = 3500$ ,  $N_w = 100$ ,  $N_T = 4500$ , and  $N_P = 1900$ . The hyperparameters are  $\rho = 0.1$ ,  $\beta = 4.1$ , and  $\alpha = 10^{-4}$ , the mean VPT is 4.9, and the longest VPT is 7.21 Lyapunov times. In Figs. 9(a) and 9(b) the reconstructed maps inside the chaotic attractors before and after the crisis are represented by red circles, respectively. Obviously, the reconstructed attractors (or the map inside attractors) fit the actual ones very well: the precisions are  $\sigma = 5.6 \times 10^{-5}$  and  $\sigma = 2.0 \times 10^{-3}$ , respectively. Moreover, the largest Lyapunov exponents of the autonomous reservoir for the cases are  $\lambda_R = 0.058$  and  $\lambda_R =$

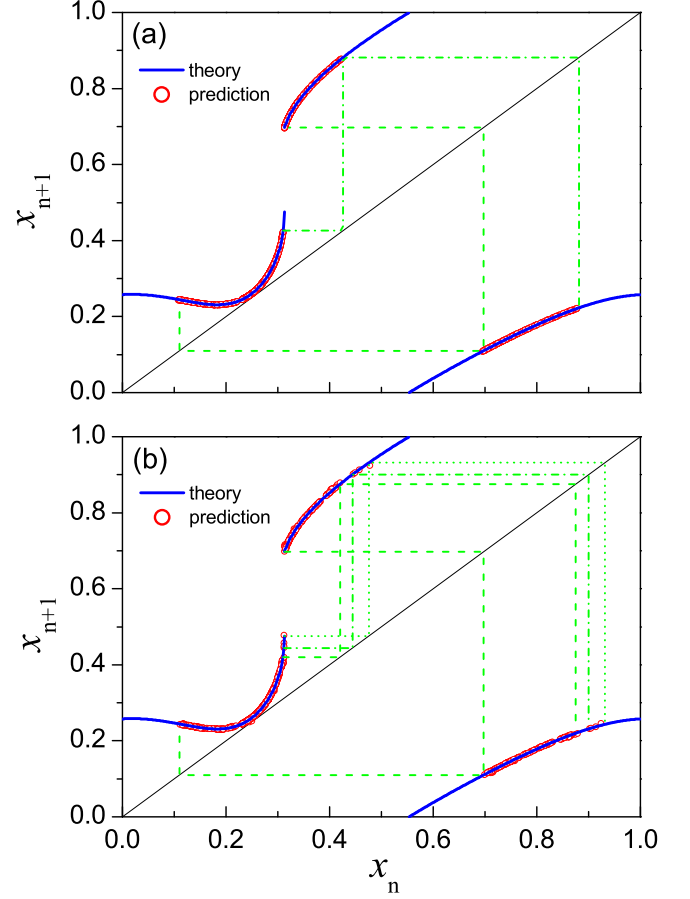


FIG. 9. Reconstructed maps (chaotic attractor) of the both noninvertible and discontinuous compound circle maps. Boundaries of the chaos bands are marked by the forward iterations of the minimum (dot-dash lines), the up border of the gap and its forward iterations (dash lines), and the down border of the gap and its forward iterations (dot lines). (a)  $U_0 = 12.57842$ ; (b)  $U_0 = 12.57950$ .

0.073, respectively, and the corresponding relative errors are 3.6% and 6.4%.

The above results imply that this RC has successfully learned the nonsmooth and discontinuous natures of the compound circle maps, and made pretty good reconstruction of the band-3 and band-5 chaotic attractors before and after the crisis, respectively. However, we have to point out that a much larger reservoir and longer training time are used in the predictions here. Nevertheless, both the mean VPT and the precision of replication are less than those in the piecewise linear discontinuous map. It can be explained by the much more complicated attractor structures. In addition, the map is nonlinear, but the activation function is piecewise linear, which also requires more network resources and data to approach the dynamics.

#### D. Lozi map

The final example is a two-dimensional nonsmooth dynamical system, the Lozi map, which reads

$$\begin{aligned} x_{n+1} &= 1 - a|x_n| + y_n, \\ y_{n+1} &= bx_n. \end{aligned} \quad (16)$$



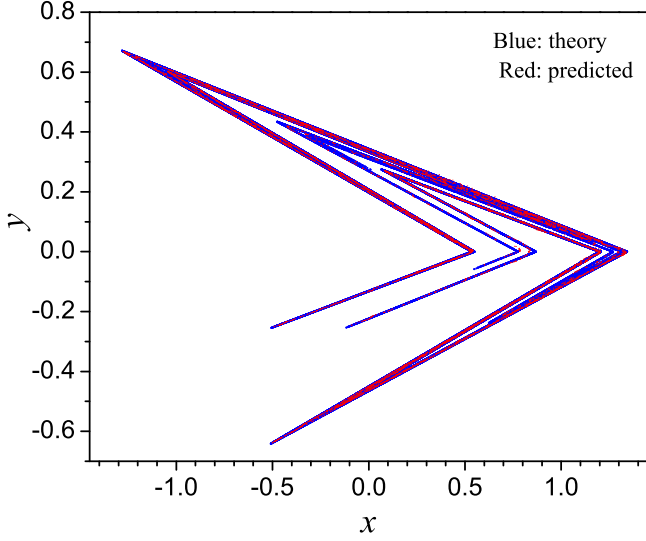


FIG. 10. Reconstructed Lozi attractor. The blue dots represent the trajectories of the actual map, and the red dots are the predictions.

This map is an area-preserving map when  $|b| = 1$ . Here we choose  $a = 1.7$  and  $b = 0.5$ , and the maximum Lyapunov exponent is  $\lambda_m = 0.47$  where the system is in the chaotic state. In the forecasting, the reservoir parameters are  $D_r = 500$ ,  $N_T = 1900$ , and  $N_W = 100$ . The optimized hyperparameters are  $\rho = 0.1$ ,  $\beta = 4.3$ , and  $\alpha = 10^{-10}$ . The numeric simulation shows that the mean valid prediction time is  $\langle T_\lambda \rangle = 6.11$ , and the longest VPT is 8.46 Lyapunov times. These data imply that the RC performs very well in the short-term prediction. A long-term prediction with  $N_p = 10000$  is plotted in Fig. 10, where the prediction precision is  $\sigma = 3.8 \times 10^{-3}$ . One can see that the trajectories (red dots) fall very closely on the actual attractor (blue dots). The largest Lyapunov exponent of the autonomous reservoir is  $\lambda_R = 0.45$ , which approximates well that of the target attractor, and the relative error is 4.3%. These results reveal that the RC with the nonsmooth activation function also performs well in predicting the two-dimensional nonsmooth dynamics.

## V. CONCLUSION

This work shows that reservoir computing with smooth activation functions does not perform well in predicting the dynamics of nonsmooth systems, especially in replicating the long-term dynamics. The reason is that the smooth hyperbolic tangent activation function cannot match the nonsmoothness of this kind of systems, and thus RC cannot capture their dynamical nature. To solve the problem, a piecewise linear activation function is proposed by linearizing the hyperbolic tangent function, which makes the activation function a continuous sigmoidal function and has nonsmooth characteristics as well. The modified RC with this activation function is a kind of promising physics-informed model for the nonsmooth dynamics. Its feasibility has been proven by the successful applications to forecast the dynamics of four nonsmooth systems with different complexity in one or two dimensions: the tent map, piecewise linear map with a gap, implicit map composed of circle maps, and Lozi map. The machine performs extremely well in the short-term prediction, producing pretty longer valid prediction times in these systems. In the long-term prediction, the perfect reconstruction of chaotic attractors is observed with higher precision. The magnitude of the precision for reconstructing chaotic attractors depends on their complexity. Larger network size and longer training time are required in replicating chaotic attractors with more complicated structures. The machine learning model here is efficient and easy to run. This work provides a primary example of research for precisely and efficiently forecasting the chaotic dynamics of nonsmooth systems.

Here we have to emphasize that the bifurcation embedment has been a hot topic in the RC community recently, and algorithms for RC with an additional channel have been proposed [45–49]. The related works are only for smooth systems. There immediately arises a question of predicting the bifurcation of nonsmooth systems. A typical characteristic of the nonsmooth systems is the border-collision bifurcation [23–25], which may result in nonsmoothness or discontinuity in the dynamics, which may be treated by introducing the piecewise linear activation function to the RC. It seems that forecasting this kind of bifurcations is straightforward by the

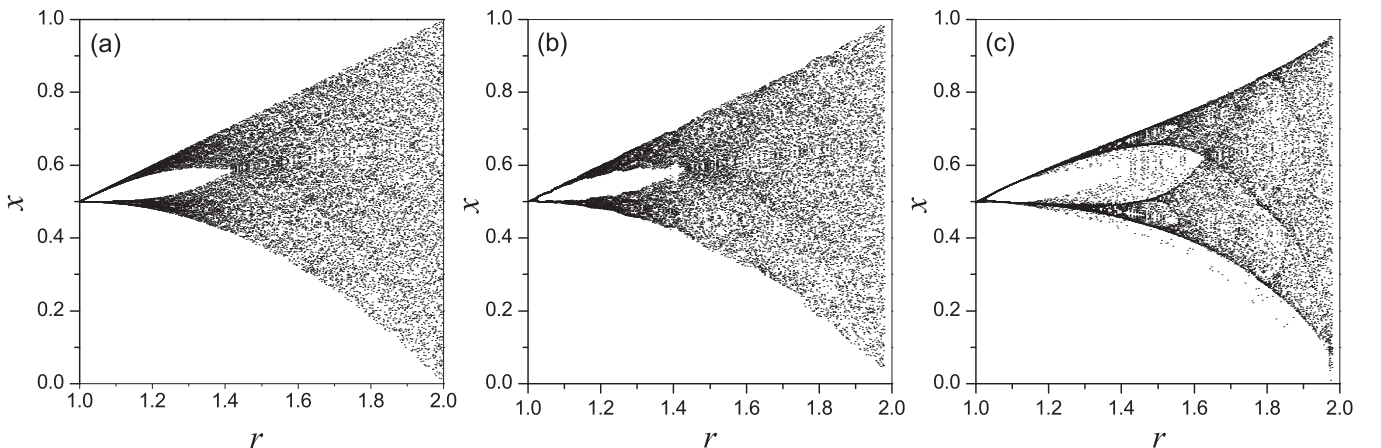


FIG. 11. Bifurcation diagram of the tent map. (a) Actual bifurcation diagram. (b) Prediction by the RC with the piecewise linear activation function proposed in this work. (c) Prediction by the RC with the hyperbolic tangent activation function.

RC with the proposed piecewise linear activation function and an additional channel for the control parameters. Indeed, our primary investigation along this route has successfully duplicated the bifurcation diagram of the tent map, as shown in Fig. 11. Obviously, the RC with the nonsmooth activation function produces a much better reconstruction of the bifurcation diagram, compared with its smooth counterpart. As for more complex nonsmooth systems, such as the piecewise linear map and compound circle maps in this work, there are complicated bifurcation structures, as shown in Figs. 6 and 8, where the coexistence of attractors frequently appears. Here the coexistence is induced by the interaction between discontinuity and noninvertibility, which is different from smooth systems. Thus, it is a bit of challenge to deal with. We have been working on this issue, which will be reported in a separate paper.

### ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under Grant No. 11975144. S.-X.Q.'s gratitude is due to Prof. Y.-C. Lai for introducing the field of machine learning.

### APPENDIX: STABILITY ANALYSIS OF THE RESERVOIR

As a driven dynamical system, the reservoir neural network requires that the reservoir driven by the same input signals asymptotically converges to the same state when starting from different initial states  $\mathbf{r}_0$  and  $\tilde{\mathbf{r}}_0$ . This means that the reservoir is stable, and thus the evolution of the internal dynamics should depend only on the input signals, but not on the initial states. This condition for the reservoir is also referred to the so-called echo-state property, which is often considered a necessary criterion for the ability of a reservoir [9,50]. The conditional Lyapunov exponents can be used to determine whether the reservoir satisfies this requirement [50]. The conventional reservoir with a hyperbolic tangent activation function can guarantee the echo-state property. In our current work, however, the activation function is piecewise linear. The

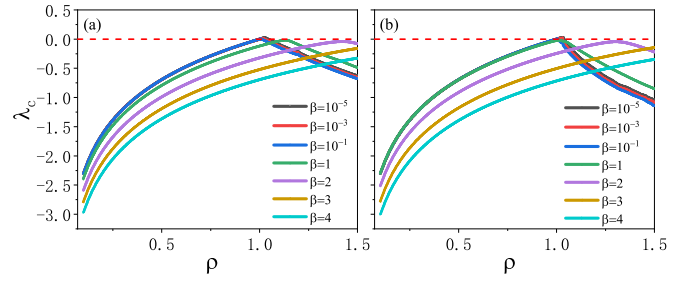


FIG. 12.  $\rho$  dependence of the largest conditional Lyapunov exponent  $\lambda_c$ . (a) For RC with hyperbolic tangent activation function  $\tanh(x)$ ; (b) for RC with the nonsmooth activation function.

echo-state property of a reservoir with this activation function and the role of hyperparameters (such as input weight scaling  $\beta$  and spectral radius  $\rho$ ) should be examined. Thus, the largest conditional Lyapunov exponents  $\lambda_c$  of RCs with this kind of activation function should be calculated. The  $\lambda_c$  are obtained by the QR decomposition algorithm when the reservoirs are driven by uniformly distributed white noise signals in the interval  $[-1, 1]$ . The results are shown in Fig. 12.

Figures 12(a) and 12(b) show the  $\rho$  dependence of the largest conditional Lyapunov exponents for RC with the hyperbolic tangent activation function and the proposed nonsmooth activation function for different  $\beta$ , respectively. Obviously, there are no significant differences between the largest conditional Lyapunov exponents of RC with two types of activation functions. In both cases, the largest conditional Lyapunov exponent  $\lambda_c$  increases to a maximum value first and then decreases with the increase of  $\rho$ . The largest Lyapunov exponent  $\lambda_c$  is negative in most parameter regimes of  $\rho$  and  $\beta$ . It is positive only when  $\beta$  is very small and  $\rho$  is around 1. These results suggest that our activation function does not significantly change the echo-state property of the reservoir. The roles of the hyperparameters are nearly the same. In this work all parameters are away from the regime of positive  $\lambda_c$ , which may guarantee the echo-state property of the reservoir. Therefore, the reservoir with the nonsmooth activation function can efficiently forecast the dynamics of the nonsmooth systems.

- [1] H. Jaeger and H. Hass, *Science* **304**, 78 (2004).
- [2] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, *Phys. Rev. Lett.* **120**, 024102 (2018).
- [3] J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. R. Hunt, M. Girvan, and E. Ott, *Chaos* **28**, 041101 (2018).
- [4] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, *Chaos* **27**, 121102 (2017).
- [5] K. Srinivasan, N. Coble, J. Hamlin, T. Antonsen, E. Ott, and M. Girvan, *Phys. Rev. Lett.* **128**, 164101 (2022).
- [6] J. Jiang and Y.-C. Lai, *Phys. Rev. Res.* **1**, 033056 (2019).
- [7] M. Rafayelyan, J. Dong, Y. Tan, F. Krzakala, and S. Gigan, *Phys. Rev. X* **10**, 041037 (2020).
- [8] L.-W. Kong, Y. Weng, B. Glaz, M. H. Haile, and Y.-C. Lai, *Chaos* **33**, 033111 (2023).
- [9] H. Jaeger, “echo state” approach to analysing and training recurrent neural networks, GMD Report 148 (German National Research Center for Information Technology, 2001).
- [10] W. Maass, T. Natschl ger, and H. Markram, *Neural Comput.* **14**, 2531 (2002).
- [11] B. Penkovsky, X. Porte, M. Jacquot, L. Larger, and D. Brunner, *Phys. Rev. Lett.* **123**, 054101 (2019).
- [12] D. Canaday, A. Griffith, and D. J. Gauthier, *Chaos* **28**, 123119 (2018).
- [13] H. Fan, J. Jiang, C. Zhang, X. Wang, and Y.-C. Lai, *Phys. Rev. Res.* **2**, 012080(R) (2020).
- [14] M. Roy, S. Mandal, C. Hens, A. Prasad, N. V. Kuznetsov, and M. D. Shrimali, *Chaos* **32**, 101104 (2022).
- [15] Z. Lu, B. R. Hunt, and E. Ott, *Chaos* **28**, 061104 (2018).

- [16] J. A. Platt, A. Wong, S. G. Clark, Randall ans Penny, and H. D. I. Abarbanel, *Chaos* **31**, 123118 (2021).
- [17] C. Zhang, S.-X. Qu, and Y.-C. Lai, *Chaos* **30**, 083114 (2020).
- [18] T. Weng, H. Yang, C. Gu, J. Zhang, and M. Small, *Phys. Rev. E* **99**, 042203 (2019).
- [19] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, *Chaos* **27**, 041102 (2017).
- [20] R. S. Zimmermann and U. Parlitz, *Chaos* **28**, 043118 (2018).
- [21] M. di Bernardo, C. J. Budd, A. R. Champneys, P. Kowalczyk, A. B. Nordmark, G. O. Tost, and P. T. Piiroinen, *SIAM Rev.* **50**, 629 (2008).
- [22] D. J. W. Simpson, *SIAM Rev.* **58**, 177 (2016).
- [23] H. E. Nusse and J. A. Yorke, *Physica D* **57**, 39 (1992).
- [24] H. E. Nusse, E. Ott, and J. A. Yorke, *Phys. Rev. E* **49**, 1073 (1994).
- [25] M. di Bernardo, C. J. Budd, and A. R. Champneys, *Phys. Rev. Lett.* **86**, 2553 (2001).
- [26] D.-R. He, B.-H. Wang, M. Bauer, S. Habip, U. Krueger, W. Martienssen, and B. Christiansen, *Physica D* **79**, 335 (1994).
- [27] B. Christiansen, D.-R. He, S. Habip, M. Bauer, U. Krueger, and W. Martienssen, *Phys. Rev. A* **45**, 8450 (1992).
- [28] S.-X. Qu, Y.-Z. Lu, L. Zhang, and D.-R. He, *Chin. Phys. B* **17**, 4418 (2008).
- [29] S.-X. Qu, S. Wu, and D.-R. He, *Phys. Rev. E* **57**, 402 (1998).
- [30] M. Bauer, S. Habip, D. R. He, and W. Martienssen, *Phys. Rev. Lett.* **68**, 1625 (1992).
- [31] Y. He, Y.-M. Jiang, Y. Shen, and D.-R. He, *Phys. Rev. E* **70**, 056213 (2004).
- [32] S.-X. Qu, B. Christiansen, and D.-R. He, *Phys. Lett. A* **201**, 413 (1995).
- [33] M. Dutta, H. E. Nusse, E. Ott, J. A. Yorke, and G. Yuan, *Phys. Rev. Lett.* **83**, 4281 (1999).
- [34] M. A. Hassouneh, E. H. Abed, and H. E. Nusse, *Phys. Rev. Lett.* **92**, 070201 (2004).
- [35] Y.-C. Lai, D.-R. He, and Y.-M. Jiang, *Phys. Rev. E* **72**, 025201(R) (2005).
- [36] W. A. S. Barbosa, A. Griffith, G. E. Rowlands, L. C. G. Govia, G. J. Ribeill, M.-H. Nguyen, T. A. Ohki, and D. J. Gauthier, *Phys. Rev. E* **104**, 045307 (2021).
- [37] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, *Proc. Natl. Acad. Sci. USA* **118**, e2101784118 (2021).
- [38] A. Choudhary, J. F. Lindner, E. G. Holliday, S. T. Miller, S. Sinha, and W. L. Ditto, *Phys. Rev. E* **101**, 062207 (2020).
- [39] C.-D. Han, B. Glaz, M. Haile, and Y.-C. Lai, *Phys. Rev. Res.* **3**, 023156 (2021).
- [40] M. Raissi, P. Perdikaris, and G. Karniadakis, *J. Comput. Phys.* **378**, 686 (2019).
- [41] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, *Nat. Rev. Phys.* **3**, 422 (2021).
- [42] A. Haluszczynski, J. Aumeier, J. Herteux, and C. R  th, *Chaos* **30**, 063136 (2020).
- [43] L. Grigoryeva and J.-P. Ortega, *Neural Netw.* **108**, 495 (2018).
- [44] G. Cybenko, *Math. Control Signals Syst.* **2**, 303 (1989).
- [45] J. Z. Kim, Z. Lu, E. Nozari, G. J. Pappas, and D. S. Bassett, *Nat. Mach. Intell.* **3**, 316 (2021).
- [46] L.-W. Kong, H.-W. Fan, C. Grebogi, and Y.-C. Lai, *Phys. Rev. Res.* **3**, 013090 (2021).
- [47] D. Patel, D. Canaday, M. Girvan, A. Pomerance, and E. Ott, *Chaos* **31**, 033149 (2021).
- [48] C. Klos, Y. F. Kalle Kossio, S. Goedeke, A. Gilra, and R.-M. Memmesheimer, *Phys. Rev. Lett.* **125**, 088103 (2020).
- [49] L. Shi, Y. Yan, H. Wang, S. Wang, and S.-X. Qu, *Phys. Rev. E* **107**, 054209 (2023).
- [50] M. Inubushi and K. Yoshimura, *Sci. Rep.* **7**, 10199 (2017).