Attractor reconstruction by machine learning EP

Cite as: Chaos **28**, 061104 (2018); https://doi.org/10.1063/1.5039508
Submitted: 08 May 2018 . Accepted: 05 June 2018 . Published Online: 22 June 2018

Zhixin Lu, Brian R. Hunt, and Edward Ott

Scilight
Highlights of the best new research in the physical sciences
LEARN MORE!
AIP Publishing

Attractor reconstruction by machine learning

Zhixin Lu,[1,a)] Brian R. Hunt,[2] and Edward Ott[3]
[1]*Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[2]*Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*
[3]*Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, Maryland 20742, USA*

(Received 8 May 2018; accepted 5 June 2018; published online 22 June 2018)

A machine-learning approach called "reservoir computing" has been used successfully for short-term prediction and attractor reconstruction of chaotic dynamical systems from time series data. We present a theoretical framework that describes conditions under which reservoir computing can create an empirical model capable of skillful short-term forecasts and accurate long-term ergodic behavior. We illustrate this theory through numerical experiments. We also argue that the theory applies to certain other machine learning methods for time series prediction. *Published by AIP Publishing*. https://doi.org/10.1063/1.5039508

A long-standing problem is prediction and analysis of data generated by a chaotic dynamical system whose equations of motion are unknown. Techniques based on delay-coordinate embedding have been successful for sufficiently low-dimensional systems. Recently, machine-learning approaches such as reservoir computing have shown promise in treating a larger class of systems. We develop a theory of how prediction with reservoir computing or related machine-learning methods can "learn" a chaotic system well enough to reconstruct the long-term dynamics of its attractor.

I. INTRODUCTION

Reservoir computing[1–4] is a machine-learning approach that has demonstrated success at a variety of tasks, including time series prediction[5–8] and inferring unmeasured variables of a dynamical system from measured variables.[9,10] In this approach, a "reservoir" is a high-dimensional, non-autonomous (driven) dynamical system, chosen independently of the task. A particular task provides an input time series, and the reservoir state as a function of time is regarded as a "raw" output time series, which is post-processed to fit the task. The post-processing function is determined, typically by linear regression, from a limited-time "training" data set consisting of the desired output time series for a given input time series.

Reservoir computing can be performed entirely in software, typically with an artificial neural network model, or with a physical reservoir; examples of the latter include a bucket of water,[11] an electronic circuit with a time delay,[12] a field-programmable gate array (FPGA),[13] an optical network of semiconductor lasers,[14] and an optic-electronic phase-delay system.[15] Other machine-learning techniques, including deep learning,[16,17] attempt to optimize internal system parameters to fit the training data; doing so requires a mathematical model of the machine-learning system. By contrast, reservoir computing does not require a model for the reservoir, nor the ability to alter the reservoir dynamics, because it seeks only to optimize the parameters of the post-processing function. The ability to use a physical reservoir as a "black box" allows for various potential advantages over other machine-learning techniques, including greatly enhanced speed.

In this article, we consider the task of predicting future measurements from a deterministic dynamical system, whose equations of motion are unknown, from limited time series data. We describe a general framework that includes the reservoir computing prediction method proposed by Jaeger and Haas.[5] With appropriate modifications, the same framework applies to other machine-learning methods for time series prediction [including a long short-term memory (LSTM) approach[18]] as we discuss further in Sec. IV. We assume the vector $\mathbf{u}(t)$ of measurements to be a function $\mathbf{h}$ of the finite-dimensional system state $\mathbf{s}(t)$

$$\mathbf{u}(t) = \mathbf{h}(\mathbf{s}(t)). \tag{1}$$

For simplicity, we assume that there is no measurement noise although our discussion below could be modified for the case that Eq. (1) is an approximation. We do not assume that $\mathbf{h}$ is invertible nor that $\mathbf{h}$ or $\mathbf{s}$ is known in practice. Training data consists of a finite time series $\{\mathbf{u}(t)\}$ of measurements. We predict future values of $\mathbf{u}(t)$ by a sequence of three steps, which we call listening, training, and predicting.

Listening consists of using the training time series as input to the reservoir, which we model as a discrete time deterministic process

$$\mathbf{r}(t+\tau) = \mathbf{f}[\mathbf{r}(t), \mathbf{u}(t)]. \tag{2}$$

Here, $\mathbf{r}(t)$ is the reservoir state, τ is a time increment, and we assume $\mathbf{f}$ to be a differentiable function. We emphasize that in practice, a formula for $\mathbf{f}$ need not be known; only its outputs are used for training and prediction. For convenience, we assume that the full reservoir state $\mathbf{r}(t)$ can be measured or computed although our arguments can be modified easily for the more general case that the reservoir output is a function of its internal state. We call Eq. (2) the "listening reservoir."

[a)] Electronic mail: zhixinlu@seas.upenn.edu.

Training consists of determining a post-processing function $\hat{\boldsymbol{\psi}}$ that, when applied to the reservoir output $\mathbf{r}(t+\tau)$, estimates the next input $\mathbf{u}(t+\tau)$. (We view $\hat{\boldsymbol{\psi}}$ as an approximation to an "ideal" post-processing function $\boldsymbol{\psi}$, to be introduced in Sec. II B.) Thus, the goal of training is to find $\hat{\boldsymbol{\psi}}$ such that $\hat{\boldsymbol{\psi}}(\mathbf{r}(t+\tau)) \approx \mathbf{u}(t+\tau)$, or equivalently

$$\hat{\boldsymbol{\psi}}(\mathbf{r}(t)) \approx \mathbf{u}(t), \tag{3}$$

for t large enough that the listening reservoir (2) has evolved beyond transient dynamics. We compute $\hat{\boldsymbol{\psi}}$ by a fitting procedure, such as linear regression, on the training time series $\{\mathbf{u}(t)\}$ and the corresponding time series $\{\mathbf{r}(t)\}$ determined from the listening reservoir (2).

Predicting then proceeds by modifying the reservoir to run autonomously with a feedback loop, replacing its input [$\mathbf{u}(t)$ in Eq. (2)] with its post-processed output from the previous time increment

$$\hat{\mathbf{r}}(t+\tau) = \mathbf{f}\left[\hat{\mathbf{r}}(t), \hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t))\right]. \tag{4}$$

We call Eq. (4) the "predicting reservoir." When initialized (from the listening reservoir state) with $\hat{\mathbf{r}}(t_0) = \mathbf{r}(t_0)$, iterating the predicting reservoir yields a time series $\{\hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t_0+\tau)), \hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t_0+2\tau)), \ldots\}$ of predictions for future measurements $\{\mathbf{u}(t_0+\tau), \mathbf{u}(t_0+2\tau), \ldots\}$. (Our notation reflects the fact that for $t > t_0$, the predicting reservoir state $\hat{\mathbf{r}}(t)$ estimates the state $\mathbf{r}(t)$ that would result from evolving the listening reservoir (2) with the future measurements.)

The reservoir prediction method we have described has been shown to produce successful short-term forecasts for a variety of dynamical systems.[5,7,8] If the system has a chaotic attractor, then, as for any imperfect model, the prediction error $\|\hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t)) - \mathbf{u}(t)\|$ cannot remain small for $t \gg t_0$. However, in some cases, the long-term time series $\{\hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t))\}$ continues to behave like the measurements from a typical trajectory on the attractor, and in this sense the predicting reservoir (4) approximately reproduces the ergodic properties of the dynamical system that generated the measurements.[8] We refer to this ability, often called attractor reconstruction, as replication of the "climate."

In this article, we develop and illustrate a theory of how reservoir prediction is able to "learn" the dynamics of a system well enough to produce both accurate short-term forecasts and accurate long-term climate. We make use of the notion of *generalized synchronization*,[19–22] which in our context means that the reservoir state $\mathbf{r}(t)$ becomes asymptotically a continuous function $\boldsymbol{\phi}$ of $\mathbf{s}(t)$, in the limit that the listening reservoir (2) is run infinitely long. In Sec. II, we argue that the following four conditions are sufficient for both short-term prediction and attractor/climate replication.

1. The listening reservoir (2) achieves generalized synchronization with the process $\{\mathbf{s}(t)\}$, so that $\mathbf{r}(t) \approx \boldsymbol{\phi}(\mathbf{s}(t))$ for a continuous function $\boldsymbol{\phi}$, within the time interval covered by the training time series.
2. The synchronization function $\boldsymbol{\phi}$ is one-to-one, or at least carries enough information about its input to recover $\mathbf{u}(t) = \mathbf{h}(\mathbf{s}(t))$ from $\boldsymbol{\phi}(\mathbf{s}(t))$.
3. Training is successful in finding a function $\hat{\boldsymbol{\psi}}$ such that Eq. (3) holds, or equivalently in view of generalized synchronization, that $\hat{\boldsymbol{\psi}}(\boldsymbol{\phi}(\mathbf{s}(t))) \approx \mathbf{h}(\mathbf{s}(t))$.
4. The attractor approached by the listening reservoir is also stable for the predicting reservoir (4).

Conditions 1–3 enable short-term prediction. Condition 4 ensures that the climate established by generalized synchronization of the listening reservoir is preserved when its input is replaced by a feedback term to form the predicting reservoir. One of the main points of Sec. II is to precisely formulate the stability condition described in condition 4.

We remark that generalized synchronization of the listening reservoir[6,10] is related to the "echo state property,"[1,23] which states that an infinite history of inputs $\{\mathbf{u}(t-\tau), \mathbf{u}(t-2\tau), \ldots\}$ uniquely determines $\mathbf{r}(t)$, subject to the condition that the trajectory $\{\mathbf{r}(t)\}$ is bounded. Indeed, if $\{\mathbf{s}(t)\}$ is a trajectory of an invertible dynamical system, then the past inputs are functions of $\mathbf{s}(t)$, so the echo state property implies that if the listening reservoir (2) has run for an infinite period of time in a bounded domain, then $\mathbf{r}(t)$ is a function of $\mathbf{s}(t)$ [though it does not imply that this function is continuous]. We believe that for the reservoir prediction method we described, it is desirable (though not strictly necessary) to have the echo state property and generalized synchronization. In Sec. II A, we show why both properties hold if the listening reservoir is uniformly contracting as a function of $\mathbf{r}$, and that we can quantify the amount of transient time it takes for the reservoir to achieve the approximation $\mathbf{r}(t) \approx \boldsymbol{\phi}(\mathbf{s}(t))$ of condition 1.

Conditions 2 and 3 are significantly more difficult to ensure *a priori*. In Sec. II B, we argue why it is plausible that these conditions can be achieved. In Secs. II C and II D, we describe the consequences of conditions 1–3 for short-term prediction and formulate more precisely the stability criterion of condition 4 that determines whether the correct attractor and climate are approximately reproduced by the long-term dynamics of the predicting reservoir (4). In Sec. II E, we describe how a model for the reservoir dynamics can be used to compute Lyapunov exponents that reflect climate stability.

In Sec. III, we give examples of short-term state and long-term climate predictions using the Lorenz equations as our input system. In addition to a case where the climate is approximated well, we show a case where the predicted climate is inaccurate, although the short-term forecast is still reasonably accurate. We compute the Lyapunov exponents of the predicting reservoir (4) and show that the transition from accurate climate to inaccurate climate corresponds to a Lyapunov exponent crossing zero. When this Lyapunov exponent is positive but close to zero, the reservoir prediction remains close to the correct climate for a transient period, and we relate the average duration of this transient to the value of the Lyapunov exponent.

II. THEORY

We consider the application of the reservoir prediction method described in the introduction to a time series $\{\mathbf{u}(t)\}$

that is a function $\mathbf{h}$ of a trajectory $\{\mathbf{s}(t)\}$ of the dynamical system

$$\mathbf{s}(t+\tau) = \mathbf{g}(\mathbf{s}(t)), \tag{5}$$

where $\mathbf{g}$ is differentiable and invertible, and we assume that $\mathbf{s}(t)$ evolves on a bounded attractor A. In preparation for training and prior to prediction, the reservoir state $\mathbf{r}(t)$ evolves according to the listening reservoir (2). The system described by Eqs. (5) and (2), coupled by Eq. (1), is often called a drive-response, skew-product, or one-way coupled system. The coupled system dynamics are illustrated by Fig. 1. We next consider the evolution of the coupled system as $t \to \infty$.

A. Listening and generalized synchronization

The goal of training can be regarded as finding a post-processing function $\hat{\boldsymbol{\psi}}$ such that $\hat{\boldsymbol{\psi}}(\mathbf{r}(t))$ is in approximate identical synchronization[20] with $\mathbf{u}(t) = \mathbf{h}(\mathbf{s}(t))$, when $\mathbf{r}(t)$ is evolved with the listening reservoir (2). The desired relationship $\mathbf{u}(t) \approx \hat{\boldsymbol{\psi}}(\mathbf{r}(t))$ can also be thought of as approximate generalized synchronization between $\mathbf{u}(t)$ [or the underlying state $\mathbf{s}(t)$] and $\mathbf{r}(t)$. The existence of such a relationship would be implied by stochastic synchronization,[19] which in our context means a one-to-one correspondence between $\mathbf{r}(t)$ and $\mathbf{s}(t)$ in the limit $t \to \infty$. However, in drive-response systems, the definition of *generalized synchronization*[21,22] requires only that the response state be asymptotically a function of the drive state: in our case, that there is a continuous function $\boldsymbol{\phi}$ such that $\mathbf{r}(t) - \boldsymbol{\phi}(\mathbf{s}(t)) \to 0$ as $t \to \infty$. The existence of such a $\boldsymbol{\phi}$ is typically easier to establish than its invertibility. Next, we describe conditions on the reservoir system $\mathbf{f}$ that guarantee generalized synchronization.

Although weaker conditions are possible, we assume uniform contraction for $\mathbf{f}$, as is often the case in practice. By *uniform contraction*, we mean that there is some $\rho < 1$ such that for all $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{u}$ we have that $|\mathbf{f}(\mathbf{r}_1, \mathbf{u}) - \mathbf{f}(\mathbf{r}_2, \mathbf{u})| < \rho|\mathbf{r}_1 - \mathbf{r}_2|$. It then follows that two trajectories $\{\mathbf{r}_1(t), \mathbf{u}(t)\}$ and $\{\mathbf{r}_2(t), \mathbf{u}(t)\}$ of (2) with the same input time series approach each other exponentially: $|\mathbf{r}_1(t) - \mathbf{r}_2(t)| \le |\mathbf{r}_1(0) - \mathbf{r}_2(0)|\rho^{t/\tau}$. Thus, for a given input time series $\{\mathbf{u}(t)\}$, the reservoir state $\mathbf{r}(t)$ is asymptotically independent of its initial state; this is essentially what Jaeger[1] called the "echo state property." Furthermore, because $\mathbf{g}$ is invertible and A is bounded, and due to the results of Hirsch, Pugh, and Shub[24,25] (a direct proof is given by Stark[26]), uniform contraction implies generalized synchronization, as defined above. (In general, the synchronization function $\boldsymbol{\phi}$ cannot be determined analytically from $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$.) A weaker form of generalized synchronization can also be guaranteed[26] from the non-uniform contraction implied by negative conditional Lyapunov exponents.

$$\begin{tikzcd}
\cdots \arrow[r, "\mathbf{g}"] & \mathbf{s}(t) \arrow[r, "\mathbf{g}"] \arrow[d, "\mathbf{h}"] & \mathbf{s}(t+\tau) \arrow[r, "\mathbf{g}"] \arrow[d, "\mathbf{h}"] & \cdots \\
\cdots \arrow[dr] & \mathbf{u}(t) \arrow[dr] & \mathbf{u}(t+\tau) \arrow[dr] & \cdots \\
\cdots \arrow[r, "\mathbf{f}"] & \mathbf{r}(t) \arrow[r, "\mathbf{f}"] & \mathbf{r}(t+\tau) \arrow[r, "\mathbf{f}"] & \cdots
\end{tikzcd}$$

FIG. 1. Drive-response system dynamics, with the drive state $\mathbf{s}(t)$ coupled to the listening reservoir state $\mathbf{r}(t)$ through the measurement vector $\mathbf{u}(t)$.

We remark that if the listening reservoir (2) is uniformly contracting, then $\mathbf{r}(t) - \boldsymbol{\phi}(\mathbf{s}(t))$ converges to zero exponentially. If the designer of the reservoir can guarantee a specific contraction rate ρ, this determines the convergence rate, so that the amount of transient time needed to make the approximation $\mathbf{r}(t) \approx \boldsymbol{\phi}(\mathbf{s}(t))$ accurate can be known in practice.

Generalized synchronization implies that the set of $(\mathbf{s}, \mathbf{r})$ such that $\mathbf{s}$ is on its attractor A and $\mathbf{r} = \boldsymbol{\phi}(\mathbf{s})$ is an attractor for the drive-response system given by Eqs. (5), (1), and (2). Below we will use the fact that this set is invariant: $\mathbf{r}(t) = \boldsymbol{\phi}(\mathbf{s}(t))$ implies $\mathbf{r}(t+\tau) = \boldsymbol{\phi}(\mathbf{s}(t+\tau))$.

B. Training

Recall that training seeks a function $\hat{\boldsymbol{\psi}}$ that predicts the current measurement vector $\mathbf{u}(t)$ from the current listening reservoir state $\mathbf{r}(t)$ [which is computed from past measurements] and that when generalized synchronization is achieved, accuracy of this prediction is equivalent to $\hat{\boldsymbol{\psi}}(\boldsymbol{\phi}(\mathbf{s}(t))) \approx \mathbf{h}(\mathbf{s}(t))$. For the rest of Sec. II, we assume that there is a function $\boldsymbol{\psi}$ defined on $\boldsymbol{\phi}(A)$ such that $\boldsymbol{\psi}(\boldsymbol{\phi}(\mathbf{s})) = \mathbf{h}(\mathbf{s})$ for all $\mathbf{s}$ in A. This assumption means that in the asymptotic limit of generalized synchronization, the listening reservoir state $\mathbf{r}(t) = \boldsymbol{\phi}(\mathbf{s}(t))$ uniquely determines $\mathbf{u}(t) = \mathbf{h}(\mathbf{s}(t))$. The goal of training can then be described as finding a function $\hat{\boldsymbol{\psi}}$ defined on the state space of the reservoir that approximates $\boldsymbol{\psi}$ on the set $\boldsymbol{\phi}(A)$. We summarize our notation in Table I.

Although the existence of $\boldsymbol{\psi}$ is not strictly necessary for the reservoir to make useful predictions, if no such $\boldsymbol{\psi}$ exists, then it seems unlikely that training can successfully achieve the desired approximation $\boldsymbol{\psi}(\boldsymbol{\phi}(\mathbf{s}(t))) \approx \mathbf{h}(\mathbf{s}(t))$, and thus

TABLE I. Summary of notation.

Dynamical system to be predicted	
$\mathbf{s}(t)$	System state
$\mathbf{g} : \mathbf{s}(t) \to \mathbf{s}(t+\tau)$	System evolution
A	Attractor for $\mathbf{s}(t)$
Measurements	
$\mathbf{u}(t)$	Measurement vector
$\mathbf{h} : \mathbf{s}(t) \to \mathbf{u}(t)$	Measurement function
Reservoir	
$\mathbf{r}(t)$	Listening reservoir state
$\mathbf{f} : [\mathbf{r}(t), \mathbf{u}(t)] \to \mathbf{r}(t+\tau)$	Listening reservoir evolution
$\hat{\mathbf{r}}(t)$	Predicting reservoir state
$\hat{\mathbf{u}}(t) = \hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t))$	Predicted measurements
$\mathbf{f} : [\hat{\mathbf{r}}(t), \hat{\mathbf{u}}(t)] \to \hat{\mathbf{r}}(t+\tau)$	Predicting reservoir evolution
Generalized synchronization	
$\boldsymbol{\phi} : \mathbf{s} \to \mathbf{r}$ for $\mathbf{s}$ in A	Synchronization function
$\boldsymbol{\psi} : \mathbf{r} \to \mathbf{u}$ for $\mathbf{r}$ in $\boldsymbol{\phi}(A)$	Ideal post-processing function
$\hat{\boldsymbol{\psi}} : \hat{\mathbf{r}}(t) \to \hat{\mathbf{u}}(t)$	Actual post-processing function

unlikely that $\mathbf{u}(t)$ can be approximated as a function of the reservoir state during either listening or predicting. The existence of $\boldsymbol{\psi}$ is guaranteed if $\boldsymbol{\phi}$ is one-to-one on A; then $\boldsymbol{\psi} = \mathbf{h} \circ \boldsymbol{\phi}^{-1}$. Furthermore, if $\mathbf{h}$ is one-to-one on A (in other words, the measurements at a given time determine the system state), then $\boldsymbol{\phi}$ must be one-to-one on A in order for $\boldsymbol{\psi}$ to exist. Thus, we propose that a goal of reservoir design should be to yield a one-to-one synchronization function $\boldsymbol{\phi}$ for a variety of input systems. In practice, having a sufficiently high-dimensional reservoir may suffice; embedding results[27,28] imply that if the dimension of the reservoir state $\mathbf{r}$ is more than twice the dimension of A, functions from A to the reservoir state space are typically one-to-one. We note that in practice, the dimension of $\mathbf{r}$ must be much larger than twice the dimension of A in order to provide a suitable basis for approximating $\boldsymbol{\psi}$, in the sense described below.

Careful consideration of conditions under which training is successful in determining an accurate approximation $\hat{\boldsymbol{\psi}}$ to $\boldsymbol{\psi}$ is beyond the scope of our theory. However, we argue that success is plausible if the training time series is sufficiently long that the trajectory $\{\mathbf{s}(t)\}$ well samples its attractor A, if the dimension of the reservoir state $\mathbf{r}(t)$ is sufficiently high, and if the dynamics of the coordinates of $\mathbf{r}(t)$ are sufficiently heterogeneous. If, for example, training uses linear regression of $\{\mathbf{u}(t)\} = \{\mathbf{h}(\mathbf{s}(t))\}$ versus $\{\mathbf{r}(t)\}$, then since $\mathbf{r}(t) \approx \boldsymbol{\phi}(\mathbf{s}(t))$, the coordinates of the vector-valued function $\boldsymbol{\phi}(\mathbf{s})$ can be thought of "basis functions";[6] training seeks a linear combination $\hat{\boldsymbol{\psi}}$ of these basis functions that approximates $\mathbf{h}(\mathbf{s})$ on A. A suitable basis for training (using a linear or nonlinear combination) is plausible if the listening reservoir yields a sufficiently large variety of responses to its input.

C. Prediction and attractor reconstruction

After training determines the post-processing function $\hat{\boldsymbol{\psi}}$, prediction proceeds by initializing $\hat{\mathbf{r}}(t_0) = \mathbf{r}(t_0)$ and evolving $\hat{\mathbf{r}}(t)$ for $t \geq t_0$ according to the predicting reservoir (4). The reservoir state $\mathbf{r}(t_0)$ is determined by evolving the listening reservoir (2) for an interval of time preceding t_0; this could be the time interval used for training, or it could be a later time interval that uses inputs $\{\mathbf{u}(t)\}$ measured after training (we call this feature "training reusability"[29]). We assume that the listening time preceding t_0 is sufficiently long to achieve generalized synchronization, so that $\hat{\mathbf{r}}(t_0) = \mathbf{r}(t_0) \approx \boldsymbol{\phi}(\mathbf{s}(t_0))$ is near $\boldsymbol{\phi}(A)$. For $t \geq t_0$, the predicted value of $\mathbf{u}(t)$ is

$$\hat{\mathbf{u}}(t) = \hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t)). \tag{6}$$

Figure 2 depicts the dynamics of the predicting reservoir (4).

Consider now the idealized scenario that our approximations are instead exact relations $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}$ on $\boldsymbol{\phi}(A)$, and $\hat{\mathbf{r}}(t_0) = \mathbf{r}(t_0) = \boldsymbol{\phi}(\mathbf{s}(t_0))$. Suppose hypothetically that the measurements $\{\mathbf{u}(t)\}$ for $t \geq t_0$ (these are the values we want to predict in practice) are available, so that we can evolve both the listening reservoir (2) depicted in Fig. 1 and the predicting reservoir (4) depicted in Fig. 2, and compare their outputs. Then we claim that the two reservoirs agree exactly: $\hat{\mathbf{r}}(t) = \mathbf{r}(t)$ and $\hat{\mathbf{u}}(t) = \mathbf{u}(t)$ for all $t \geq t_0$. First notice that $\hat{\mathbf{u}}(t_0) = \hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t_0)) = \boldsymbol{\psi}(\boldsymbol{\phi}(\mathbf{s}(t_0))) = \mathbf{h}(\mathbf{s}(t_0)) = \mathbf{u}(t_0)$. Then $\hat{\mathbf{r}}(t_0 + \tau) = \mathbf{f}[\hat{\mathbf{r}}(t_0), \hat{\mathbf{u}}(t_0)] = \mathbf{f}[\mathbf{r}(t_0), \mathbf{u}(t_0)] = \mathbf{r}(t_0 + \tau)$, and $\mathbf{r}(t_0 + \tau) = \boldsymbol{\phi}(\mathbf{s}(t_0 + \tau))$ due to generalized synchronization. Similarly, $\hat{\mathbf{u}}(t_0 + \tau)$ then equals $\mathbf{u}(t_0 + \tau)$, so $\hat{\mathbf{r}}(t_0 + 2\tau) = \mathbf{r}(t_0 + 2\tau) = \boldsymbol{\phi}(\mathbf{s}(t_0 + 2\tau))$, etc. This agreement between the trajectories also shows that $\boldsymbol{\phi}(A)$ is an invariant set for the idealized predicting reservoir

$$\mathbf{r}(t + \tau) = \mathbf{f}[\mathbf{r}(t), \boldsymbol{\psi}(\mathbf{r}(t))], \tag{7}$$

and that its dynamics, observed through $\boldsymbol{\psi}$, are equivalent to the dynamics of A observed through $\mathbf{h}$.

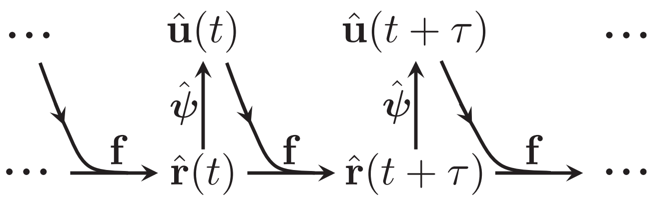

···
$\hat{\mathbf{u}}(t)$
$\hat{\mathbf{u}}(t+\tau)$
···
$\hat{\boldsymbol{\psi}}$
$\hat{\boldsymbol{\psi}}$
···
$\mathbf{f}$
$\hat{\mathbf{r}}(t)$
$\mathbf{f}$
$\hat{\mathbf{r}}(t+\tau)$
$\mathbf{f}$
···

FIG. 2. Predicting reservoir dynamics, with the listening reservoir input $\mathbf{u}(t)$ replaced by the estimate $\hat{\mathbf{u}}(t)$ determined from the predicting reservoir state $\hat{\mathbf{r}}(t)$.

Thus, if the time series $\{\mathbf{u}(t)\}$ of measurements has enough information to reconstruct the attractor A, then we can regard $\boldsymbol{\phi}(A)$ and the idealized predicting reservoir (7) as an exact reconstruction of A and its dynamics. When the approximation $\hat{\boldsymbol{\psi}} \approx \boldsymbol{\psi}$ is not exact on $\boldsymbol{\phi}(A)$, the actual predicting reservoir (4) is still initialized near $\boldsymbol{\phi}(A)$, but $\boldsymbol{\phi}(A)$ is only approximately invariant. The better the approximation, the more accurate the predictions $\hat{\mathbf{u}}(t) \approx \mathbf{u}(t)$ will be, at least in the short term. However, if the system (5) that generates the measurements $\{\mathbf{u}(t)\}$ is chaotic, the prediction error $\|\hat{\mathbf{u}}(t) - \mathbf{u}(t)\|$ will typically grow exponentially as t increases.

Nonetheless, it remains possible that $\hat{\mathbf{u}}(t)$ will maintain a climate similar to $\mathbf{u}(t)$ in the long term. This will happen if (and practically speaking, only if) the predicting reservoir trajectory $\{\hat{\mathbf{r}}(t)\}$ remains close to $\boldsymbol{\phi}(A)$ for all time, and its attractor has a similar climate to that of the idealized predicting reservoir on $\boldsymbol{\phi}(A)$. In this sense, climate replication (attractor reconstruction) relies on both state-space stability and structural stability of the predicting reservoir near the idealized reconstructed attractor $\boldsymbol{\phi}(A)$.

Structural stability is difficult to ensure rigorously, but in practice small perturbations of the dynamics near an attractor tend to yield small perturbations to the climate. Thus, we argue that climate replication is likely if $\boldsymbol{\phi}(A)$, which according to our assumptions is invariant for the idealized predicting reservoir, is also attracting, in the sense described below.

D. Stability and Lyapunov exponents

Recall that generalized synchronization implies that the set $\boldsymbol{\phi}(A)$ is attracting for the listening reservoir (2), when driven by $\mathbf{u}(t) = \mathbf{h}(\mathbf{s}(t))$ where $\mathbf{s}(t)$ evolves on A. Whether $\boldsymbol{\phi}(A)$ is attracting for the predicting reservoir is complicated by the fact that it is invariant only in the idealized case $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi}$ and that $\boldsymbol{\psi}$ is defined only on $\boldsymbol{\phi}(A)$, so that the idealized predicting reservoir (7) is also defined only on $\boldsymbol{\phi}(A)$. For its stability to be well-defined, the domain of $\boldsymbol{\psi}$ must be extended to a neighborhood of $\boldsymbol{\phi}(A)$, and whether $\boldsymbol{\phi}(A)$ is attracting depends on how the extension is chosen.

Thus, the suitability of the empirically determined function $\hat{\boldsymbol{\psi}}$ for climate prediction depends not only on how well it approximates $\boldsymbol{\psi}$ on $\boldsymbol{\phi}(A)$ but also on how it behaves near $\boldsymbol{\phi}(A)$. For a particular $\hat{\boldsymbol{\psi}}$, we consider hypothetically a particular extension of $\boldsymbol{\psi}$ such that $\hat{\boldsymbol{\psi}} \approx \boldsymbol{\psi}$ near $\boldsymbol{\phi}(A)$. This extension gives the idealized predicting reservoir a full set of Lyapunov exponents on $\boldsymbol{\phi}(A)$, some of which correspond to infinitesimal perturbations tangent to $\boldsymbol{\phi}(A)$ and some of which correspond to infinitesimal perturbations transverse to $\boldsymbol{\phi}(A)$. Then $\boldsymbol{\phi}(A)$ is attracting if the transverse Lyapunov exponents are all negative, and is unstable if there is a positive transverse Lyapunov exponent.

If the generalized synchronization function $\boldsymbol{\phi}$ is one-to-one and differentiable, then the tangential Lyapunov exponents of the system (5) on A are reproduced as the tangential Lyapunov exponents of the idealized predicting reservoir on $\boldsymbol{\phi}(A)$. Generalized synchronization does not always yield a differentiable $\boldsymbol{\phi}$,[26,30] but even when differentiability cannot be guaranteed, it is possible in practice to reproduce much of the Lyapunov spectrum of A, including negative Lyapunov exponents in some cases, with a predicting reservoir.[8]

We remark that unlike the conditional Lyapunov exponents for a drive-response system (such as the listening reservoir), which correspond to perturbations of the response system state, for the predicting reservoir it is not clear in advance which perturbations correspond to transverse Lyapunov exponents. However, in a numerical experiment where the equations for the driving system (5) and the reservoir are known, the existence or absence of a positive transverse Lyapunov exponent can be inferred by computing all of the positive Lyapunov exponents of the predicting reservoir and eliminating those that are Lyapunov exponents of A.

E. Computation of Lyapunov exponents

We now describe how to estimate the Lyapunov exponents of the idealized predicting reservoir (7) on $\boldsymbol{\phi}(A)$, for a particular extension of $\boldsymbol{\psi}$ to a neighborhood of $\boldsymbol{\phi}(A)$, from its empirical approximation $\hat{\boldsymbol{\psi}}$. To do so, we assume that we have a formula for **f**, so that we can compute its Jacobian matrix. (We emphasize that we estimate the Lyapunov exponents in order to corroborate the theory we have presented; their computation, and a formula for **f**, are not needed for the reservoir prediction method we have described.) If climate replication is successful, we can simply generate a long trajectory of the predicting reservoir (4) and use it to compute the Lyapunov exponents of the trajectory.[8] However, this trajectory cannot be expected to remain close to $\boldsymbol{\phi}(A)$ if the set is unstable. Nonetheless, if we have a sufficiently long time time series $\{\mathbf{u}(t)\}$ of measurements, we can estimate the Lyapunov exponents of $\boldsymbol{\phi}(A)$, whether or not it is stable, as follows.

First, we use the time series $\{\mathbf{u}(t)\}$ to generate a trajectory $\{\mathbf{r}(t)\}$ of the listening reservoir (2); as we have argued, $\mathbf{r}(t)$ will approach $\boldsymbol{\phi}(A)$ under the conditions for generalized synchronization. Then along this trajectory, which is an approximate trajectory for the predicting reservoir, we compute Lyapunov exponents using the Jacobian matrix of the predicting reservoir (4).

III. NUMERICAL EXPERIMENTS

In this section, we give examples of short-term state and long-term climate predictions for the Lorenz system,[31] with standard parameter values that yield chaotic trajectories

$$\begin{aligned} dx/dt &= 10(y-x),\\ dy/dt &= x(28-z)-y, \\ dz/dt &= xy - 8z/3. \end{aligned} \tag{8}$$

We consider the case where the measurement function **h** is the identity, so that $\mathbf{u}(t) = \mathbf{s}(t) = [x(t), y(t), z(t)]^T$. For the reservoir, we use an artificial neural network similar to the one used by Jaeger and Haas;[5] our listening reservoir [a continuous-time version of Eq. (2)] evolves according to

$$\frac{d}{dt}\mathbf{r}(t) = \gamma[-\mathbf{r}(t) + \tanh(\mathbf{M}\mathbf{r}(t) + \sigma\mathbf{W}_{\text{in}}\mathbf{u}(t))], \tag{9}$$

where $\mathbf{r}$ is an N-dimensional vector, γ is a scalar, and $\mathbf{M}$ is an adjacency matrix representing internal network connections. The matrix $\sigma\mathbf{W}_{\text{in}}$ consists of "input weights"; in our numerical results, we will fix $\mathbf{W}_{\text{in}}$ and vary the scalar input strength σ. The vector function tanh is computed by applying the scalar hyperbolic tangent to each coordinate of its input vector. We compute trajectories of both the Lorenz and reservoir systems using the 4th order Runge-Kutta method with time step $\tau = 0.001$. We will show cases where climate replication (attractor reconstruction) succeeds and where it fails, and compare the results with Lyapunov exponents we compute for the predicting reservoir.

We consider post-processing functions of the form $\hat{\boldsymbol{\psi}}(\mathbf{r}) = \mathbf{W}_{\text{out}}\mathbf{q}(\mathbf{r})$, where $\mathbf{q}(\mathbf{r})$ is the $2N$-dimensional vector consisting of the N coordinates of $\mathbf{r}$ followed by their squares, and the "output weight" matrix $\mathbf{W}_{\text{out}}$ is determined by a linear regression procedure described below. The listening reservoir (9) and the post-processing function are illustrated as an input-output system in Fig. 3. The goal of training is that the post-processed output $\mathbf{W}_{\text{out}}\mathbf{q}(\mathbf{r}(t+\tau))$

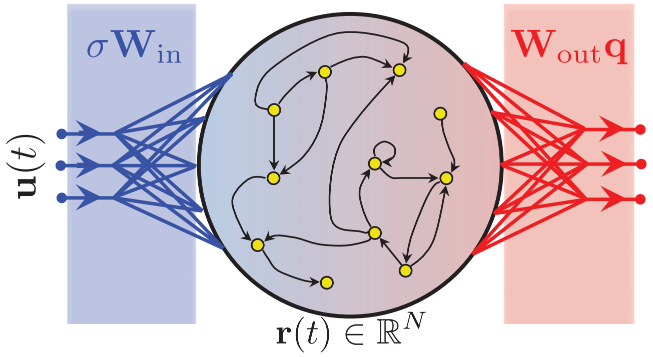

$\sigma\mathbf{W}_{\text{in}}$
$\mathbf{W}_{\text{out}}\mathbf{q}$
$\mathbf{u}(t)$
$\mathbf{r}(t) \in \mathbb{R}^N$

FIG. 3. Listening reservoir based on an artificial neural network with N neurons. The input vector $\mathbf{u}(t) \in \mathbb{R}^3$ is mapped to the reservoir state space $\mathbb{R}^N$ by the input weight matrix $\sigma\mathbf{W}_{\text{in}}$, and the resulting reservoir state is mapped to $\mathbb{R}^3$ by the post-processing function $\hat{\boldsymbol{\psi}} = \mathbf{W}_{\text{out}}\mathbf{q}$.

based on input up to time t estimates the subsequent input $\mathbf{u}(t+\tau)$. Once $\mathbf{W}_{\text{out}}$ is determined, the external input can be replaced in a feedback loop by the post-processed output to form the predicting reservoir, as depicted in Fig. 4. The predicting reservoir evolves according to

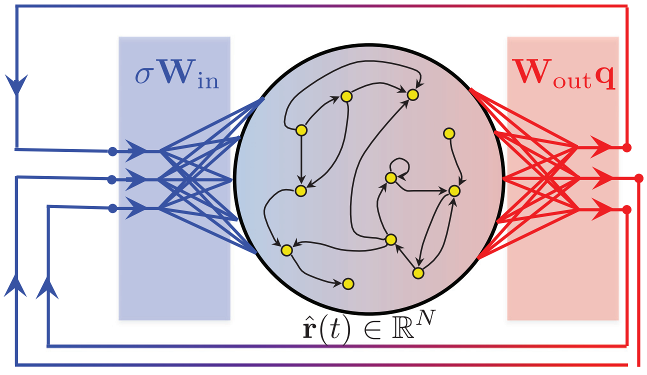
$$\frac{d}{dt}\hat{\mathbf{r}}(t) = \gamma[-\hat{\mathbf{r}}(t) + \tanh(M\hat{r}(t) + \sigma\mathbf{W}_{\text{in}}\mathbf{W}_{\text{out}}\mathbf{q}(\hat{\mathbf{r}}(t))], \quad (10)$$

and the predicted value of $\mathbf{u}(t)$ is $\hat{\mathbf{u}}(t) = \hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t)) = \mathbf{W}_{\text{out}}\mathbf{q}(\hat{\mathbf{r}}(t))$.

Details of our reservoir implementation are as follows. The reservoir dimension is $N = 2000$, and we use $\gamma = 10$. The N-by-N adjacency matrix $\mathbf{M}$ is chosen randomly with sparse Erdös-Renyi connectivity and spectral radius 0.9; specifically, each element is chosen independently to be nonzero with a probability of 0.02, nonzero elements are chosen uniformly between –1 and 1, and the resulting matrix is rescaled so that the magnitude of its largest eigenvalue is 0.9. The N-by-3 matrix $\mathbf{W}_{\text{in}}$ is chosen randomly so that each row has one non-zero element, chosen uniformly between –1 and 1. We evolve the Lorenz system and the listening reservoir (9) from time $t = -100$ to $t = 60$, and we discard 100 time units of transient evolution, so that training is based on $\mathbf{u}(t)$ and $\mathbf{r}(t)$ for $0 \leq t \leq 60$. For training, we constrain the 3-by-2N matrix $\mathbf{W}_{\text{out}}$ to have only $3N$ nonzero elements, namely, the first N elements of its first two rows, and the first $N/2$ and last $N/2$ elements of its third row. (Thus, we fit the x and y coordinates of the Lorenz state with linear functions of $\mathbf{r}$, and the z coordinate with a linear combination of the first $N/2$ coordinates of $\mathbf{r}$ and the squares of the second $N/2$ coordinates; for the Lorenz system, this is advantageous over using a purely linear function of $\mathbf{r}$.[8]) Subject to this constraint, we select $\mathbf{W}_{\text{out}}$ so as to minimize the error function

$$\sum_{k=1}^{3000} ||\mathbf{W}_{\text{out}}\mathbf{q}(\mathbf{r}(0.02k)) - \mathbf{u}(0.02k)||^2 + \beta||\mathbf{W}_{\text{out}}||^2; \quad (11)$$

here we have coarsely sampled the training data every 0.02 time units in order to reduce the amount of computation required by the regression. The second term in the error function modifies ordinary linear least-squares regression in order to discourage overfitting; this modification is often called ridge regression or Tikhonov regularization. Below, we will show results with regularization parameter $\beta = 10^{-6}$ and with $\beta = 0$ (no regularization). We begin prediction by initializing $\hat{\mathbf{r}}(T) = \mathbf{r}(T)$ and evolving the predicting reservoir (10), where $T = 60$ is the end of the listening and training periods.



$\sigma\mathbf{W}_{\text{in}}$
$\mathbf{W}_{\text{out}}\mathbf{q}$
$\hat{\mathbf{r}}(t) \in \mathbb{R}^N$

FIG. 4. The predicting reservoir replaces the external input of the listening reservoir with the post-processed reservoir output. The time increment τ in our discussion represents the amount of time for information to travel once around the feedback loop.

In Fig. 5, we show the actual $z(t)$ from a trajectory of the Lorenz system, and predictions $\hat{z}(t)$ from two reservoirs that are identical except for their input strength parameter values [$\sigma = 0.012$ for Fig. 5(a) and $\sigma = 0.014$ for Fig. 5(b)]. Each reservoir is trained with the same Lorenz trajectory and with regularization parameter $\beta = 10^{-6}$. Both reservoirs predict the short-term future similarly well, but for larger values of the prediction time $t - T$, only the second prediction continues with a Lorenz-like climate. We compare the two climate predictions over a longer time period in Fig. 6, which shows Poincaré return maps of successive $z(t)$ maxima. In Fig. 6(a), the red dots (showing the reservoir prediction) initially are near the blue dots (representing the Lorenz attractor), but eventually the red dots approach a period two orbit, indicated by the arrows. The large distance of the upper left arrow from the blue dots indicates that this period two orbit for the reservoir is not on the Lorenz attractor. In contrast, the red dots in Fig. 6(b) remain near the blue dots at all times, indicating that the reservoir replicates the climate in the long term.

Based on the arguments in Sec. II A, we hypothesize that for both $\sigma = 0.012$ and $\sigma = 0.014$, the listening reservoir (9) evolves toward a set $\boldsymbol{\phi}_\sigma(A)$, where A is the Lorenz attractor and $\boldsymbol{\phi}_\sigma$ is a generalized synchronization function. Our choice of spectral radius 0.9 for the adjacency matrix $\mathbf{M}$ is consistent with common practice in reservoir computing,[32]

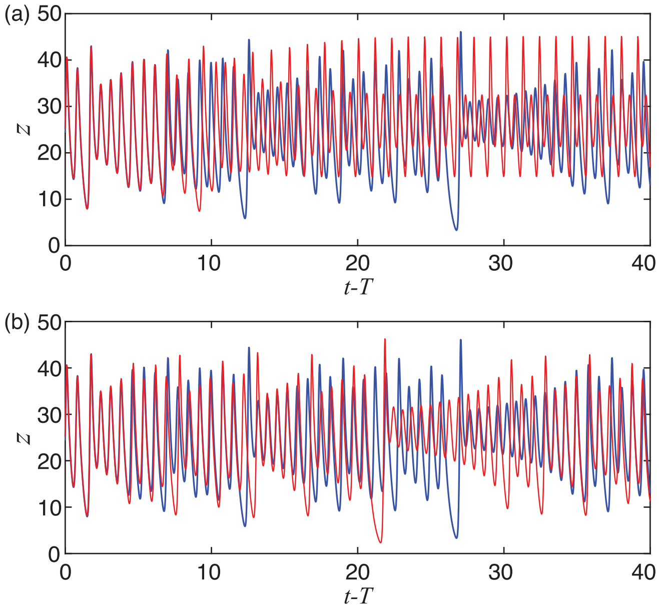


FIG. 5. Predicted (red) and actual (blue) $z(t)$ for a chaotic Lorenz system trajectory, using the same randomly generated reservoir with different input strengths $\sigma = 0.012$ [panel (a)] and $\sigma = 0.014$ [panel (b)]. Both predictions remain well correlated with the actual trajectory for roughly 10 time units. After decorrelation, the first prediction approaches a periodic orbit, whereas the second prediction appears to continue with a climate similar to that of the actual trajectory.

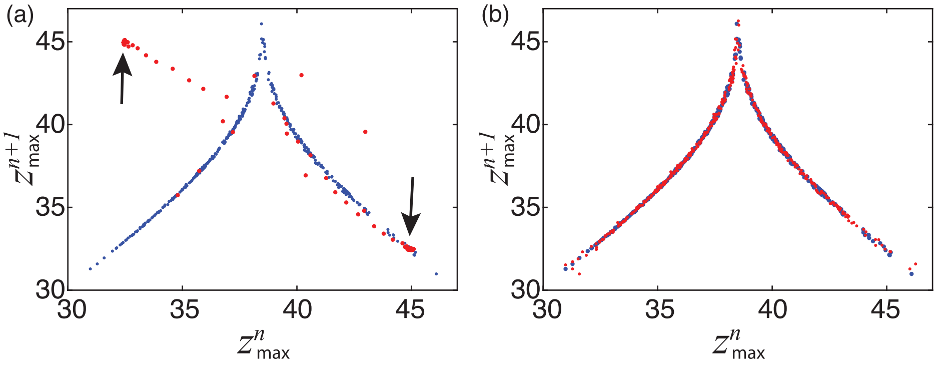


FIG. 6. Poincaré return map of successive local maxima of $z(t)$ for the actual (blue) and predicted (red) trajectories for $t - T$ from 0 to 300, using the same Lorenz trajectory and reservoir as Fig. 5, again with $\sigma = 0.012$ [panel (a)] and $\sigma = 0.014$ [panel (b)]. Here $z_{\max}^{n}$ represents the nth local maximum of $z(t)$. The first prediction approaches a period two orbit (indicated by the arrows) that is not on the Lorenz attractor, whereas the second prediction remains close to the Lorenz attractor.

although it does not guarantee uniform contraction for the listening reservoir.[23] However, it does guarantee that the eigenvalues of the Jacobian matrix of the right side of (9), evaluated at $\mathbf{r} = \mathbf{u} = 0$, have real parts at most $\gamma(-1 + 0.9) = 10(-0.1) = -1$. This suggests an asymptotic contraction rate of -1 or faster for the listening reservoir, and that after discarding 100 transient time units, $\mathbf{r}(t)$ is extremely close to $\boldsymbol{\phi}_\sigma(A)$ for $t \geq 0$.

Based on the arguments in Sec. II C, we hypothesize that the set $\boldsymbol{\phi}_\sigma(A)$ is approximately invariant for the predicting reservoir (10). Based on the results in Figs. 5 and 6, we hypothesize further that for $\sigma = 0.014$, there is an attracting invariant set for the predicting reservoir near $\boldsymbol{\phi}_\sigma(A)$, but that between $\sigma = 0.014$ and $\sigma = 0.012$, there is a bifurcation that causes this invariant set either to become unstable or to be destroyed entirely. To corroborate this hypothesis, we compute the Lyapunov exponents of the predicting reservoir for an approximate trajectory on $\boldsymbol{\phi}_\sigma(A)$, as described in Sec. II E.

Figure 7 shows the three largest Lyapunov exponents of the predicting reservoir (10) as the input strength σ varies from 0.004 to 0.02. We do not change the matrices $\mathbf{M}$ and $\mathbf{W}_{\text{in}}$, but for each value of σ, we perform a separate training (with $\beta = 10^{-6}$ as before), resulting in a different output weight matrix $\mathbf{W}_{\text{out}}$. The exponents colored red and blue approximate the positive and zero Lyapunov exponents of the Lorenz attractor A (the approximation is closest for $\sigma \geq 0.01$). Reproduction of the positive exponent of A in the reservoir dynamics on $\boldsymbol{\phi}_\sigma(A)$ is a necessary consequence of successful attractor reconstruction and does not indicate instability of $\boldsymbol{\phi}_\sigma(A)$ to transverse perturbations. The exponent colored green estimates the largest of the transverse Lyapunov exponents described in Sec. II D. This exponent passes through zero, indicating a bifurcation, at $\sigma \approx 0.013$.

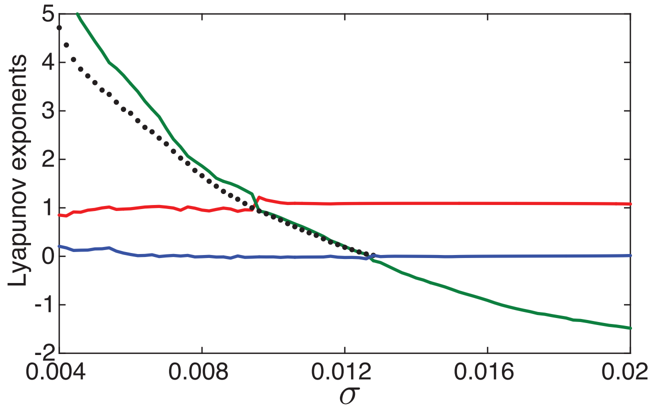


FIG. 7. The three largest Lyapunov exponents of the predicting reservoir (10) on the invariant set $\boldsymbol{\phi}_\sigma(A)$ for the listening reservoir (9), as a function of the input strength σ, for the same reservoir as Figs. 5 and 6. Two exponents that are approximately constant as a function of σ, which approximate the two largest Lyapunov exponents of the Lorenz attractor, are colored red and blue; the more variable exponent, which we call the transverse Lyapunov exponent and which determines climate stability, is colored green. For values of σ for which we detect divergence from the Lorenz climate, we graph with a black dot the observed divergence rate λ^*, computed as described in the text.

Next, we compare the change in stability indicated by the computed transient Lyapunov exponent to a more direct computation indicating success or failure of climate replication. To detect when the prediction $\hat{\mathbf{u}}(t) = \hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t))$ of the Lorenz state diverges from the true Lorenz attractor, we let $\Delta(t)$ be the Euclidean distance between the vector field $d\hat{\mathbf{u}}/dt$ implied by the predicting reservoir and the vector field (right-hand side) of the Lorenz system (8), evaluated at $[x, y, z]^T = \hat{\mathbf{u}}(t)$. [We calculate the reservoir-implied vector field by the chain rule $d\hat{\mathbf{u}}/dt = D\hat{\boldsymbol{\psi}}(\hat{\mathbf{r}}(t))d\hat{\mathbf{r}}/dt$, where $D\hat{\boldsymbol{\psi}}$ is the Jacobian matrix of $\hat{\boldsymbol{\psi}} = \mathbf{W}_{\text{out}}\mathbf{q}$, and $d\hat{\mathbf{r}}/dt$ is given by Eq. (10).] For each value of σ depicted in Fig. 7, we calculate the vector field discrepancy $\Delta(t)$ for the prediction time period $t \geq T$. If $\Delta(t)$ does not exceed a threshold value 20 for a duration of 800 time units, we consider the climate to be approximately reproduced. (Our threshold value 20 is small compared to the typical magnitude of the Lorenz vector field.) Otherwise, we say that the prediction has "escaped" from the Lorenz attractor. In Fig. 7, we show a black dot at each value of σ for which we detect escape; these values are the same as those for which the computed transverse Lyapunov exponent is positive. The height of each black dot represents an observed divergence rate λ^*, computed as follows.

When we detect escape for a particular value of σ, we reinitialize the predicting reservoir (10) using $\hat{\mathbf{r}}(t_0) = \mathbf{r}(t_0)$ for 1000 different values of $t_0 \geq T$, where the values of $\mathbf{r}(t_0)$ are determined by continuing to run the listening reservoir (9) for $t \geq T$. For each t_0, we evolve the predicting reservoir until the first time t_1 for which $\Delta(t_1) \geq 20$, or until $t_1 - t_0 = 800$, whichever comes first. If divergence from the attractor is governed by Lyapunov exponent λ, we should have $\Delta(t_1) \approx \Delta(t_0)\exp(\lambda(t_1 - t_0))$ in a certain average sense. We compute the observed exponential divergence rate $\lambda^* = \langle \ln[\Delta(t_1)/\Delta(t_0)]\rangle / \langle t_1 - t_0\rangle$, where the angle brackets represent an average over the 1000 values of t_0. The computed values of λ^* are shown as black dots in Fig. 7. The approximate agreement of

λ^* with the green curve (especially for $0.01 \le \sigma \le 0.013$) demonstrates that the computed transverse Lyapunov exponent reflects divergence of predictions from the Lorenz attractor.

To illustrate the correspondence between the computed transverse Lyapunov exponent and the observed divergence rates in a case where their dependence on σ is more complicated, we show in Fig. 8 the analogue of Fig. 7 in a case where no regularization ($\beta = 0$) is used in the training. Again, we see precise correspondence between detected failure of climate replication (presence of a black dot) and positive values of the transverse Lyapunov exponent (green curve), and good agreement with the observed divergence rates for these values of σ. In this case, there are two bifurcations, one near $\sigma = 0.12$ and one near $\sigma = 0.16$.

We remark that when we use regularization ($\beta = 10^{-6}$) in the training, we do not observe as complicated a dependence of the computed transverse Lyapunov exponent on the input strength σ as in Fig. 8. Instead, the computed transverse Lyapunov exponent is typically negative and slowly varying across a wide range of σ values, for which climate replication is successful. In Fig. 9, we use the transverse Lyapunov exponent computation, averaged over 10 different randomly generated reservoirs, to give a quantitative illustration of the advantage of regularization. When regularization is used, the negative means and small standard deviations of the computed transverse Lyapunov exponent indicate robust climate stability over the entire range $0.05 \le \sigma \le 0.5$. (By contrast, Figs. 5–7 depicted values of $\sigma \le 0.02$.) With no regularization, the means are larger and more variable, indicating less stability and greater sensitivity to the value of σ, and the standard deviations are significantly larger, indicating lack of robustness from one random reservoir realization to another.

IV. CONCLUSIONS AND DISCUSSION

We presented in Sec. II a partial explanation for how reservoir computing prediction is able to reconstruct the attractor (replicate the climate) for a chaotic process from limited time series data. We argued that the reservoir dynamics (2) can be designed so that during the listening period on which training is based, the reservoir state $\mathbf{r}(t)$ is approximately a continuous function $\boldsymbol{\phi}$ of the state $\mathbf{s}(t)$ of the chaotic process. This property, called generalized synchronization, is closely related to the echo state property for reservoir computing. We showed that both properties hold if the listening reservoir (2) is uniformly contracting as a function of the reservoir state; other criteria for these properties have also been identified.[23,26,32]

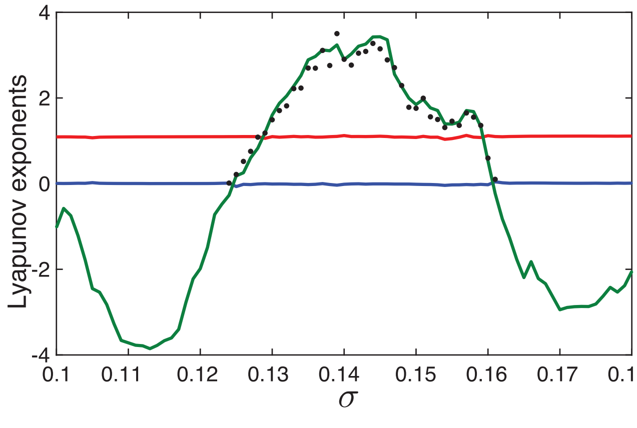


FIG. 8. The three largest Lyapunov exponents of the predicting reservoir (10), and the estimated divergence rate λ^*, as a function of σ, using the same color scheme as Fig. 7. Here we use a different randomly generated reservoir than in Fig. 7, and no regularization ($\beta = 0$) in the training.

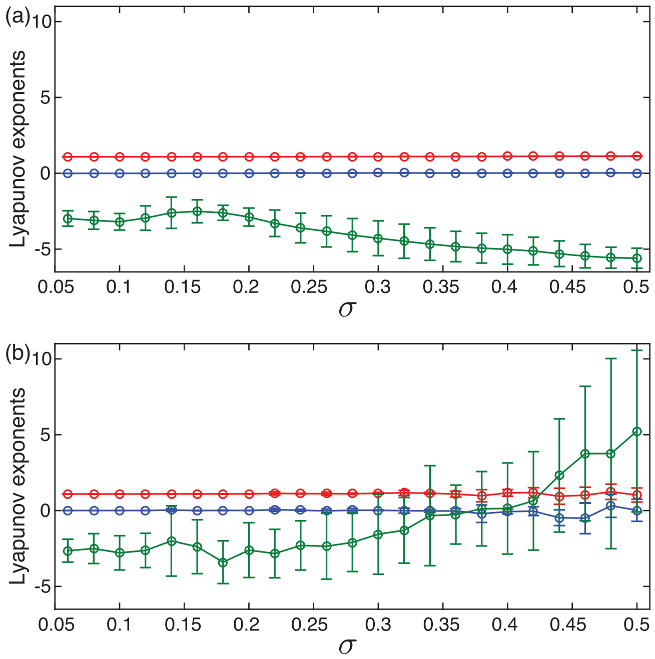


FIG. 9. The means and standard deviations of three largest Lyapunov exponents for the same 10 randomly generated reservoirs trained with regularization parameter $\beta = 10^{-6}$ [panel (a)] and with $\beta = 0$ [panel (b)]. Again, the red and blue curves approximate the two largest exponents of the Lorenz attractor, and the green curve is the computed transverse Lyapunov exponent.

Ideally, the synchronization function $\boldsymbol{\phi}$ should be one-to-one in order to recover the process dynamics from the reservoir dynamics. Investigation of conditions that can guarantee $\boldsymbol{\phi}$ to be one-to-one could help guide reservoir design. However, even in the absence of a guarantee, we noted that embedding results suggest that $\boldsymbol{\phi}$ is likely to be one-to-one if the reservoir state space is sufficiently high-dimensional compared with dimensionality of the chaotic process.

Practically speaking, a necessary condition for climate replication is that training should be successful in approximately recovering the measured state $\mathbf{u}(t) = \mathbf{h}(\mathbf{s}(t))$ from the reservoir state $\mathbf{r}(t)$; this depends on the amount of training data available and the method of regression used, among other things. We did not address the theoretical aspects of training, but we argued that success is plausible if the reservoir is sufficiently high-dimensional and heterogeneous to yield a large variety of basis functions for the regression.

We showed that in the limit that the approximations we described are exact, the predicting reservoir (4) exactly predicts future values of $\mathbf{u}(t)$. Thus, accurate approximations yield commensurately accurate short-term forecasts. Long-term climate replication depends on stability of the predicting reservoir dynamics with respect to perturbations produced by the approximations. We discussed how to estimate Lyapunov exponents for the predicting reservoir in numerical experiments, whether or not the desired climate is stable. We emphasize that our computation of Lyapunov exponents was intended to illustrate our theory and that the method we described requires measurements $\{\mathbf{u}(t)\}$ over a long time period to maintain the desired climate. If one's goal is to estimate the Lyapunov exponents of the process that produced $\{\mathbf{u}(t)\}$ from a limited amount of data, one should seek parameters of the predicting reservoir that replicate the climate and simply compute the Lyapunov exponents of the resulting trajectory.[8]

In Sec. III, we gave examples of climate replication successes and failures, and showed how they correspond to the Lyapunov exponents we computed. We emphasize that the results and the ranges of σ we displayed were selected to illustrate and analyze failures that can occur with inadequate input strength (Figs. 5–7) or without regularization (Fig. 8) in the training. With regularization, we are able to obtain robust climate replication [indicated by Fig. 9(a)] over a wide range of input strengths.

We remark that for simplicity, our theory considered discrete-time reservoir dynamics. Discrete time is the appropriate way to model software reservoirs, but physical reservoirs typically are better modeled by continuous time. With appropriate modifications, our theory applies to the continuous-time case. The prediction time increment τ used in the training should be the amount of time information takes to traverse the feedback loop depicted in Fig. 4. However, with a physical reservoir, careful calibration of the sampled training data may be necessary to meet the goal of predicting $\mathbf{u}(t+\tau)$ based on the listening reservoir's response to input up to time t, in part because τ is a property of the predicting reservoir and not of the listening reservoir.

Finally, we argue that in addition to reservoir computing, the theory we presented in Sec. II applies to some other machine learning methods for time series prediction. The essential features a prediction method needs for our theory to apply are: (1) that the method maintains an internal state, or "memory," that depends on the sequence of inputs it receives during training; (2) that it is trained to predict a short time increment ahead, after receiving the input time series for a relatively long time interval; and (3) that it is used to predict farther into the future by iterating its incremental forecasts through a feedback loop. These features are present, for example, in prediction using the FORCE method for training reservoirs[33] and in recent work using long short-term memory (LSTM) networks for prediction.[18] For methods that (unlike reservoir computing) train parameters that affect the internal state in the absence of feedback, our theory applies if we take the function $\mathbf{f}$ in Eq. (2) to represent the update rule for the internal state $\mathbf{r}$ after training has selected parameter values. Although our description of how training arrives at the pair of functions $(\mathbf{f}, \hat{\psi})$ was specific to reservoir computing, our discussion of how these functions can be used with Eqs. (4) and (6) for prediction and attractor reconstruction are independent of which machine-learning method is used to determine the functions.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of grants from ARO (W911NF-12-1-0101) and DARPA. We thank Michelle Girvan, Jaideep Pathak, Sarthak Chandra, Daniel Gauthier, and Daniel Canaday for their input.

[1] H. Jaeger, GMD Technical Report No. 148 (German National Research Center for Information Technology, 2001).

[2] W. Maass, T. Natschläger, and H. Markram, Neural Comput. **14**, 2531 (2002).

[3] H. Jaeger, Scholarpedia **2**, 2330 (2007).

[4] M. Lukoševičius and H. Jaeger, Comput. Sci. Rev. **3**, 127 (2009).

[5] H. Jaeger and H. Haas, Science **304**, 78 (2004).

[6] U. Parlitz and A. Hornstein, Chaos Complexity Lett. **1**, 135 (2005).

[7] F. Wyffels and B. Schrauwen, Neurocomputing **73**, 1958 (2010).

[8] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, Chaos **27**, 121102 (2017).

[9] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, Chaos **27**, 041102 (2017).

[10] R. S. Zimmermann and U. Parlitz, Chaos **28**, 043118 (2018).

[11] C. Fernando and S. Sojakka, in *European Conference on Artificial Life* (Springer, 2003), pp. 588–597.

[12] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Nat. Commun. **2**, 468 (2011).

[13] N. D. Haynes, M. C. Soriano, D. P. Rosin, I. Fischer, and D. J. Gauthier, Phys. Rev. E **91**, 020801 (2015).

[14] D. Brunner, S. Reitzenstein, and I. Fischer, in *IEEE International Conference on Rebooting Computing (ICRC)* (IEEE, 2016), pp. 1–2.

[15] L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, Phys. Rev. X **7**, 011015 (2017).

[16] Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).

[18] P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and P. Koumoutsakos, Proc. R. Soc. A **474**, 20170844 (2018).

[19] V. Afraimovich, N. Verichev, and M. I. Rabinovich, Radiophys. Quantum Electron. **29**, 795 (1986).

[20] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **64**, 821 (1990).

[21] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. Abarbanel, Phys. Rev. E **51**, 980 (1995).

[22] L. Kocarev and U. Parlitz, Phys. Rev. Lett. **76**, 1816 (1996).

[23] I. B. Yildiz, H. Jaeger, and S. J. Kiebel, Neural Networks **35**, 1 (2012).

[24] M. Hirsch and C. Pugh, "Global analysis," in *Proceedings of Symposia in Pure Mathematics* (American Mathematical Society, 1970), Vol. 14, pp. 133–164.

[25] M. Hirsch and C. Pugh, *Invariant Manifolds*, Lecture Notes in Mathematics No. 583 (Springer-Verlag, 1977).

[26] J. Stark, Physica D **109**, 163 (1997).

[27] H. Whitney, Ann. Math. **37**, 645 (1936).

[28] T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).

[29] J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. R. Hunt, M. Girvan, and E. Ott, Chaos **28**, 041101 (2018).

[30] B. R. Hunt, E. Ott, and J. A. Yorke, Phys. Rev. E **55**, 4029 (1997).

[31] E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

[32] K. Caluwaerts, F. Wyffels, S. Dieleman, and B. Schrauewn, in *2013 International Joint Conference on Neural Networks* (IEEE, 2014), pp. 1–6.

[33] D. Sussillo and L. F. Abbott, Neuron **63**, 544 (2009).