



OPEN

Reservoir computing with generalized readout based on generalized synchronization

Akane Ohkubo¹ & Masanobu Inubushi^{1,2,3✉}

Reservoir computing is a machine learning framework that exploits nonlinear dynamics, exhibiting significant computational capabilities. One of the defining characteristics of reservoir computing is that only linear output, given by a linear combination of reservoir variables, is trained. Inspired by recent mathematical studies of generalized synchronization, we propose a novel reservoir computing framework with a *generalized* readout, including a nonlinear combination of reservoir variables. Learning prediction tasks can be formulated as an approximation problem of a target map that provides true prediction values. Analysis of the map suggests an interpretation that the linear readout corresponds to a linearization of the map, and further that the generalized readout corresponds to a higher-order approximation of the map. Numerical study shows that introducing a generalized readout, corresponding to the quadratic and cubic approximation of the map, leads to a significant improvement in accuracy and an unexpected enhancement in robustness in the short- and long-term prediction of Lorenz and Rössler chaos. Towards applications of physical reservoir computing, we particularly focus on how the generalized readout effectively exploits low-dimensional reservoir dynamics.

Keywords Reservoir computing, Generalized synchronization, Echo state property

Reservoir computing (RC) is a machine learning framework that exploits dynamical systems and has remarkable computational capabilities^{1–3}. For example, RC using random networks, called echo state networks (ESNs), can efficiently predict chaotic time series⁴. Adding closed-loop makes an RC system autonomous and capable of replicating chaotic attractors, which are utilized to estimate Lyapunov exponents⁵. Furthermore, recent studies have shown that such ‘autonomous’ RC systems can reproduce true dynamical properties more accurately than those computed from limited training data and extrapolate true dynamical structures such as bifurcation outside the training data^{6–8}. Another branch of research, physical RC, harnesses various physical dynamics and demonstrates high information processing capability^{9–13,14}.

Why does RC work so well with untrained random networks and physical systems? This is a central open problem in RC research, and more broadly in machine learning and neuroscience. Partial answers to this problem have been provided using dynamical systems theory^{7,15–17}. In particular, Grigoryeva, Hart, and Ortega¹⁷ rigorously proved the existence of a continuously differentiable synchronization map under certain conditions and explicitly showed what the RC learns when predicting chaotic dynamics. In other words, they provided a formal expression of a map, which is h , as explained later in Eq. (8), that RC approximates for prediction. Hara and Kokubu⁷ uncovered a key mathematical structure for learning with RC, i.e. a smooth conjugacy between target and reservoir dynamics based on observations from the numerical study of the logistic map.

Inspired by these seminal studies^{7,17}, we propose a novel method of RC with a *generalized* readout. Based on generalized synchronization, the Taylor expansion of the map h , the Eq. (10), may give an interpretation of the conventional RC as a linearization of h . Moreover, it implies that the computational capabilities of RC with a generalized readout are superior to those of conventional RC. Remark that a specific type of nonlinear readout, such as $\sum_i W_i(r_t)_i^2$ in the notation introduced later, has already been used in previous research^{5,18,19}. We emphasize that our theoretical framework *comprehensively explains* the reason why these nonlinear readouts are so effective, and, furthermore, presents a new direction of research, including *cubic* readouts, bridging rigorous mathematics^{7,17} with practical applications.

Indeed, numerical studies on Lorenz and Rössler chaos prediction as a benchmark problem strongly support this; i.e., for both short- and long-term predictions, we reveal the significant computational capabilities of RC

¹Department of Applied Mathematics, Tokyo University of Science, Shinjuku, Tokyo 162-8601, Japan. ²Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK. ³Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan. ✉email: inubushi@rs.tus.ac.jp

with generalized readout compared to conventional RC. Moreover, for long-term prediction, the autonomous RC system with generalized readout acquires notable robustness, in contrast to the lack of robustness of conventional RC.

Formulation

Conventional RC

Here we briefly sketch the method of conventional RC. Let us consider the target input $x_t \in \mathbb{R}^K$, the target output $y_t \in \mathbb{R}^K$ vectors ($t \in \mathbb{Z}$), and their sequence $\{x_t, y_t\}$. The goal is to construct a machine that, given an input x_t , produces an output \hat{y}_t that approximates the target output y_t , i.e. $\hat{y}_t \simeq y_t$, by using the training data $\{x_t, y_t\}$.

The machine consists of *reservoir* variables, $r_t \in [-1, 1]^N$, whose dynamics are determined by a map $F : [-1, 1]^N \times \mathbb{R}^K \rightarrow [-1, 1]^N$ and the input x_t as follows

$$r_t = F(r_{t-1}, x_t). \quad (1)$$

The output \hat{y}_t is determined by the *readout weight* matrix $W \in M_{K,N}$ and the output bias $b \in \mathbb{R}^K$ as

$$\hat{y}_t = W r_t + b. \quad (2)$$

The readout weight matrix W^* is determined such that $\hat{y}_t \simeq y_t$:

$$W^* := \arg \min_W \langle \|y_t - \hat{y}_t\|^2 \rangle_T = \arg \min_W \langle \|y_t - W r_t\|^2 \rangle_T, \quad (3)$$

as usual the method of least squares, where $\langle \cdot \rangle_T$ denotes the long-term average. For simplicity, the regularization term is omitted in the formulation, but it is used, as explained in the following numerical study.

Synchronizations

Common-Signal-Induced Synchronization (CSIS), or equivalently the Echo State Property (ESP) in the context of RC, is a key property required for reservoir dynamics determined by the map F . For a given (common) input signal $\{x_t\}_{t \geq 1}$ and arbitrary initial reservoir states r_0, \tilde{r}_0 ($r_0 \neq \tilde{r}_0$), we say that CSIS occurs if the reservoir states converge to a unique state that depends only on the sequence of the input signal, i.e.

$$\lim_{t \rightarrow +\infty} \|r_t - \tilde{r}_t\| = 0, \quad (4)$$

where these dynamics are determined by

$$r_t = F(r_{t-1}, x_t), \quad \tilde{r}_t = F(\tilde{r}_{t-1}, x_t). \quad (5)$$

The occurrence of CSIS can be characterized by the conditional Lyapunov exponent^{15,16}.

In this paper, we only study the input signal $\{x_t\}_{t \geq 1}$ is generated by another dynamical system, referred to as the *target* dynamical system determined by a nonlinear map $\phi : \mathbb{R}^K \rightarrow \mathbb{R}^K$,

$$x_t = \phi^t(x_0), \quad (6)$$

where $x_0 \in \mathbb{R}^K$ denotes the initial point of the target dynamics and $\phi^t = \phi \circ \dots \circ \phi$ is the t times composition of ϕ . It is more general to formulate the observation function of the target dynamics, ω , as in Grigoryeva et al.¹⁷, however, we do not consider it for simplicity.

Note that, if CSIS occurs, the asymptotic states of the reservoir dynamics $\{r_t\}$ are uniquely determined by the target dynamics $\{x_t\}$ after the transient period. This correspondence is referred to as *generalized synchronization*, and denoted by

$$r_t = f(x_t), \quad (7)$$

where $f : \mathbb{R}^K \rightarrow \mathbb{R}^N$ is the generalized synchronization map. See Fig. 1 a for an illustration of . Grigoryeva et al.¹⁷ proved the existence and the differentiability of the map under certain conditions.

Generalized readout

Let us consider the inverse of the map exists, and then, $x_t = f^{-1}(r_t)$. In that case, for instance, τ -ahead prediction task of the target dynamics can be expressed by

$$y_t = x_{t+\tau} = \phi^\tau(x_t) = \phi^\tau(f^{-1}(r_t)) =: h(r_t) \quad (8)$$

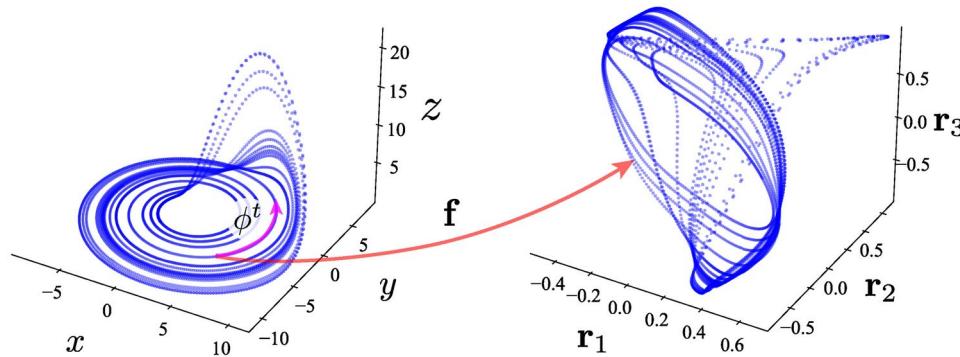


Fig. 1. An illustration of the target and reservoir dynamics in phase space. As an example of the target dynamical system, the Rössler attractor is shown in the left panel. The right panel shows the projection of the reservoir dynamics driven by the Rössler dynamics onto the subspace spanned by the first three variables, i.e. r_1 , r_2 , r_3 . The red arrow depicts the schematic of the generalized synchronization map f .

as a function of the reservoir state r_t . Therefore, predicting τ -ahead target dynamics with RC is mathematically equivalent to the functional approximation of the map $h : \mathbb{R}^N \rightarrow \mathbb{R}^K$. This indicates that the conventional RC may be viewed as a linearisation of the map h (see the Eq. (2)), i.e.,

$$h(r_t) \simeq Wr_t + b. \quad (9)$$

However, the map h is not linear in general, since it is the composition of the nonlinear maps of ϕ^τ and f^{-1} . Here we assume h is sufficiently smooth, and consider the Taylor expansion of h ,

$$h(r) = h(0) + \nabla h(0)r + \frac{1}{2}r^T H(0)r + O(\|r\|^3), \quad (10)$$

where $\nabla h \in M_{K,N}$ and $H \in M_{K,N,N}$ denote the Jacobian $[\nabla h]_{ij} = \frac{\partial h_i}{\partial r_j}$ ($i = 1, \dots, K$, $j = 1, \dots, N$) and the Hessian $[H]_{ijk} = \frac{\partial^2 h_i}{\partial r_j \partial r_k}$ ($i = 1, \dots, K$, $j, k = 1, \dots, N$), respectively. This gives an interpretation of the conventional RC; that is, the output bias vector b and the readout weight matrix W are used to approximate the first two terms in the Taylor expansion as

$$b \simeq h(0), \quad W \simeq \nabla h(0), \quad (11)$$

with the approximation error of $O(\|r\|^2)$. In this sense, the conventional RC may be understood as a linear approximation of the target map h .

In this paper, we propose to utilize the *nonlinear* combination of the reservoir variables for the approximation of higher order terms in the Taylor expansion. In other words, our method, referred to as RC with *generalized readout*, approximates the general term in the Taylor expansion beyond the linear term. Taking into account up to second order, we include the quadratic form $r_t^T W^Q r_t$ into the output as

$$\hat{y}_t = Wr_t + r_t^T W^Q r_t \quad (12)$$

so that the readout weight tensor, $W^Q \in M_{K,N,N}$, approximates the Hessian term as

$$W^Q \simeq \frac{1}{2}H(0). \quad (13)$$

with the approximation error of $O(\|r\|^3)$. This corresponds to the quadratic approximation of the map h . As the conventional RC, the readout weights W^* and W^{Q*} are determined such that

$$(W^*, W^{Q*}) := \arg \min_{W, W^Q} \langle \|y_t - \hat{y}_t\|^2 \rangle_T \quad (14)$$

$$= \arg \min_{W, W^Q} \langle \|y_t - Wr_t - r_t^T W^Q r_t\|^2 \rangle_T, \quad (15)$$

which we refer to as *quadratic-form* RC (QRC).

Note that learning in our method is *linear* with respect to the weights, W and W^Q , which result in again least squares, and therefore, retains the simplicity of the conventional RC, i.e., the low computational cost and

guaranteed optimality. Furthermore, the output of our method is *nonlinear* with respect to the reservoir variables as $(r_t)_i(r_t)_j$, which leads to a greater variety of approximations to the functional relationship between r on $h(r)$. Moreover, it is natural to expect that including higher terms, beyond QRC, will give a better approximation, and indeed, we show in the following numerical experiments that this is true at least up to the third order (see Fig. 7).

Numerical study of QRC and beyond

We numerically show that the RC with generalized readout is superior to the conventional RC for the prediction task, and that the closed-loop long-term prediction using the QRC provides better performance with unexpected robustness.

Here, we use the Echo State Network (ESN) as a reservoir, i.e., the map F is given by $F(r, x) = \tanh(Ar + Bx)$, where the component-wise application of \tanh is employed as the activation function, $A \in M_{N,N}$, and $B \in M_{N,K}$ are the random matrices. The elements of the random matrices A and B are sampled independently and identically from a uniform random distribution over the interval $[-\sigma_A, \sigma_A]$ and $[-\sigma_B, \sigma_B]$, respectively, and we use the Ridge regression at the training phase with the regularization parameter β . Hyperparameter optimization was performed for each ESN and each task, and the detailed results are given in the Supplementary Information.

The number of training parameters, denoted by M in the following numerical study, is summarised here for later use (Fig. 3b). Henceforth, we refer to the conventional ESN as linear ESN (\mathcal{L} -ESN) and the quadratic-form ESN as \mathcal{Q} -ESN. Concerning \mathcal{L} -ESN, in accordance with previous mathematical studies^{6–8}, we employ a simple architecture without leaking rate, the special structure of the adjacency matrix, output bias, and so on, where simply $M = K \times N$ for K outputs. As for \mathcal{Q} -ESN, using the symmetry of $W^{\mathcal{Q}}$, i.e., $W_{ijk}^{\mathcal{Q}} = W_{ikj}^{\mathcal{Q}}$, we have

$$M = K \times \left(N + \frac{N \times (N+1)}{2} \right) \text{ for } K \text{ outputs.}$$

The target dynamical system is determined by the Lorenz equations; $\dot{x} = 10(y - x)$, $\dot{y} = 28x - y - xz$, $\dot{z} = xy - 8/3z$, where its time- τ map gives ϕ of the target dynamics (6) with $K = L = 3$, calculated using the fourth-order Runge-Kutta method with $\Delta t = 0.02$. As another example of dynamical systems, we show the prediction results for the Rössler system in the Supplementary Information.

Short-term prediction (open-loop)

First, we study the short-term prediction, and in particular, $\tau = 0.2$ ahead prediction of the Lorenz chaos; hence, when the input is $x_t = \phi^t(x_0)$, the target output is $y_t = \phi(x_t) = x_{t+1}$. Figure 2 shows the prediction results with the ESN size $N = 10$. Note that while we also use $N = 10$ in the long-term prediction later, which is quite small compared to the commonly used one, as a reference, $N = 300$ is used in⁵. The left panels of Fig. 2a are the time series of the target signals, i.e., y_t , depicted by the grey dashed lines, and those of the predictions, i.e., \hat{y}_t , depicted by the red solid lines. The predictions by the \mathcal{L} -ESN and the \mathcal{Q} -ESN are shown in Fig. 2a and b, respectively.

Although there is a discrepancy between the target signal $\{y(t)\}$ and the prediction by the \mathcal{L} -ESN $\{\hat{y}(t)\}$, it is difficult to distinguish between the target signal $\{y(t)\}$ and the prediction by the \mathcal{Q} -ESN $\{\hat{y}(t)\}$, i.e., the \mathcal{Q} -ESN provides a more accurate prediction than the \mathcal{L} -ESN. The right panels of Fig. 2 show the phase space structures of the orbit $\{\hat{y}(t)\}$, corresponding to the time series on the left panels. The phase space structures of the orbit $\{\hat{y}(t)\}$ by the \mathcal{L} -ESN are far from those of the true Lorenz attractor; however, the \mathcal{Q} -ESN can predict the orbit whose phase space structure is qualitatively the same as the true one, the butterfly wing shape. In summary, the \mathcal{Q} -ESN is more accurate in short-term predictive ability than the \mathcal{L} -ESN when $N = 10$.

To quantitatively compare the predictive ability of the \mathcal{L} -ESN and the \mathcal{Q} -ESN, we plot the root mean square errors (RMSE), $\sqrt{\langle \|y_t - \hat{y}_t\|^2 \rangle_T}$, in Fig. 3a for the ESN size $5 \leq N \leq 50$. The values of the RMSE over the 20 different realizations for each case are shown to investigate the dependence of the random number realizations used for the matrices A and B . Here, the red and blue circles represent the RMSE given by the \mathcal{L} -ESN and the \mathcal{Q} -ESN, respectively. While the RMSE values typically decrease with increasing ESN size, there is a huge gap between the RMSE values of the \mathcal{L} -ESN and the \mathcal{Q} -ESN, i.e., the RMSE of the \mathcal{Q} -ESN is significantly lower than that of the \mathcal{L} -ESN.

Figure 3a shows the comparison of the RMSE for the \mathcal{L} - and \mathcal{Q} -ESNs for the same network size N ; however, the number of training parameters M differs between \mathcal{L} - and \mathcal{Q} -ESNs. Fig. 3b is the same as Fig. 3a, but the horizontal axis is $M \in \{60, 195, 405, 690\}$. The RMSE values are almost the same for the \mathcal{L} - and \mathcal{Q} -ESN. Note that, in this comparison, the network sizes of the \mathcal{L} - and \mathcal{Q} -ESN are not the same. For example, in the case of $M = 690$, while the network size of the \mathcal{L} -ESN is $N = 230$, that of the \mathcal{Q} -ESN is $N = 20$. The smaller network size of the \mathcal{Q} -ESN results in the larger scatter; however, even when the number of training parameters for the comparison is fixed, the best result achieved by the \mathcal{Q} -ESN is almost the same or remarkably better than that by the \mathcal{L} -ESN, e.g., the case of $M = 690$.

Long-term prediction (closed-loop)

For the long-term prediction, the both ESNs are trained for the τ -ahead prediction task where $\tau = 0.02$ of the Lorenz chaos. Again, the target output is $y_t = \phi(x_t) = x_{t+1}$. The output from the ESN is denoted by $\hat{y}_t = \hat{h}(r_t)$, where the output function is $\hat{h}(r_t) = Wr_t$ for the \mathcal{L} -ESN and $\hat{h}(r_t) = Wr_t + r_t^T W^{\mathcal{Q}} r_t$ for the \mathcal{Q} -ESN. After training, we obtain $\hat{y}_t \simeq y_t = x_{t+1}$. In the next step, we employ $r_{t+1} = F(r_t, \hat{y}_t) = F(r_t, \hat{h}(r_t)) =: G(r_t)$ instead of (1), where the map $G : \mathbb{R}^N \rightarrow \mathbb{R}^N$ determines the *autonomous* dynamical system in the reservoir state space. This closed-loop method using the ESN, which we call the *autonomous* ESN for short, not only provides long-term prediction, but also has the surprising ability to reconstruct the target dynamics determined

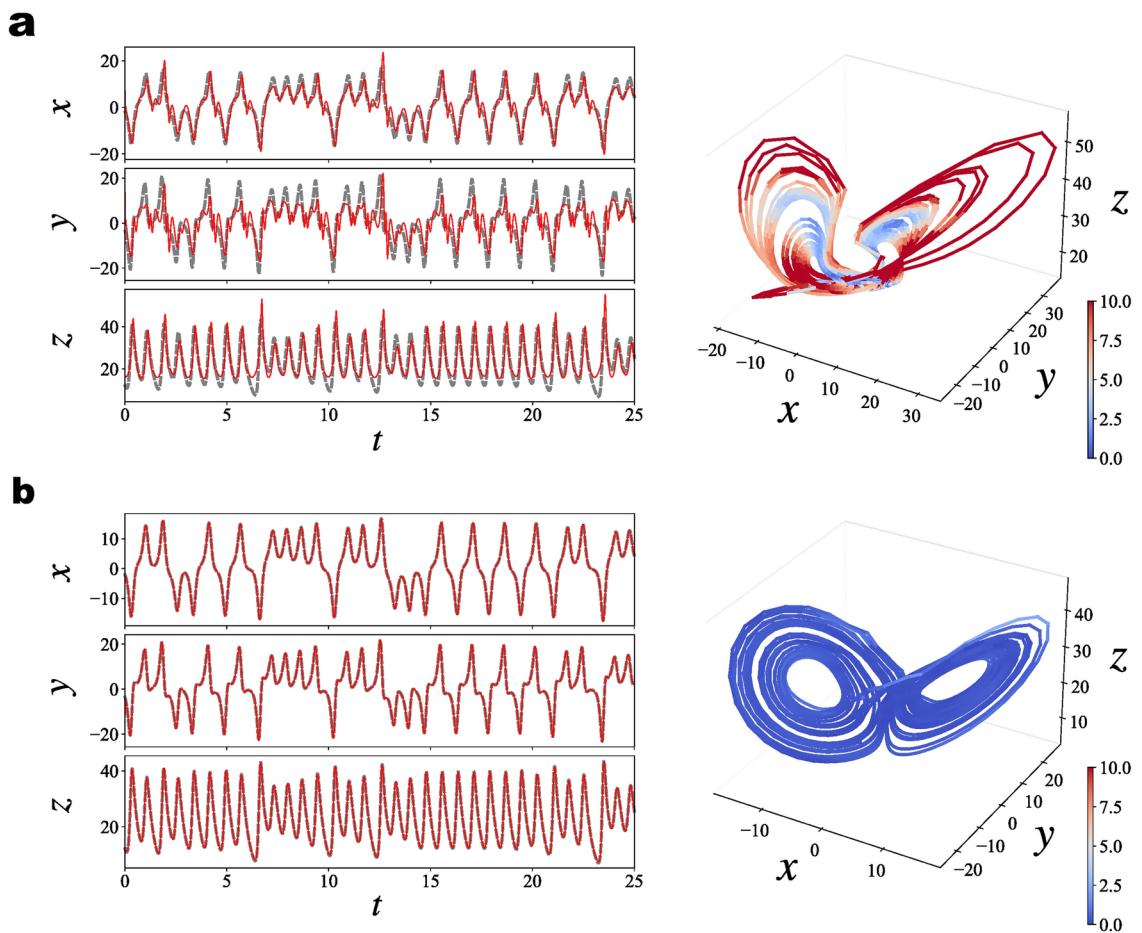


Fig. 2. Short-term prediction (open-loop). The panels (a) and (b) show the results using \mathcal{L} - and \mathcal{Q} -ESN, respectively. The left and right panels show the time series of the target (grey dashed) and prediction (red solid) and the phase space structures of the orbits. The colors represent the local error of the prediction, $\|\mathbf{y}_t - \hat{\mathbf{y}}_t\|$.

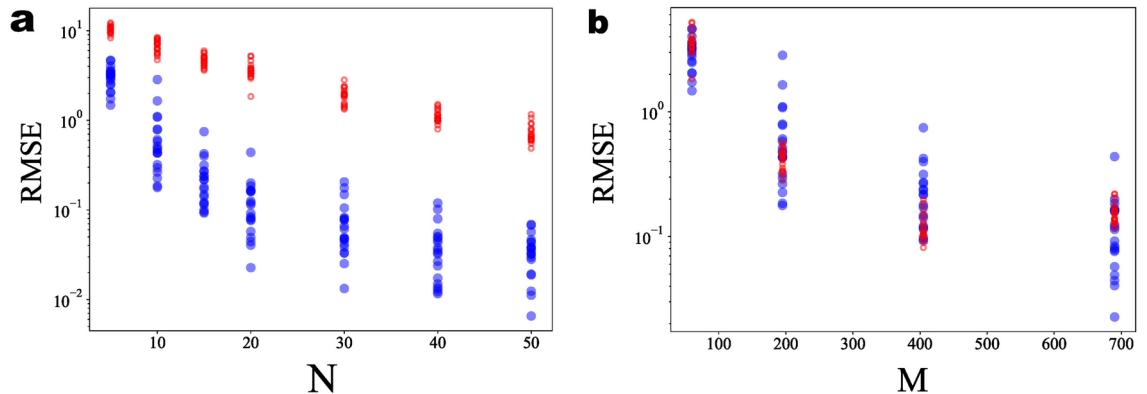


Fig. 3. Short-term prediction (open-loop). (a) The root mean square error (RMSE), $\sqrt{\langle \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|^2 \rangle_T}$, over the size of the network N . The red open and blue solid circles show the RMSE using \mathcal{L} - and \mathcal{Q} -ESN, respectively. (b) The same as (a), but the horizontal axis shows the number of trained parameters

by ϕ , as mentioned in the introduction. Here we fix $N = 10$ and examine the effect of varying the functional form $\hat{h}(\mathbf{r}_t)$, i.e. using the \mathcal{L} -ESN or the \mathcal{Q} -ESN, on these abilities.

First, Fig. 4 shows the results of long-term prediction using the autonomous \mathcal{L} -ESN. The results depend on the realization of the random matrices A and B , we show two cases of different random numbers in Fig. 4a,

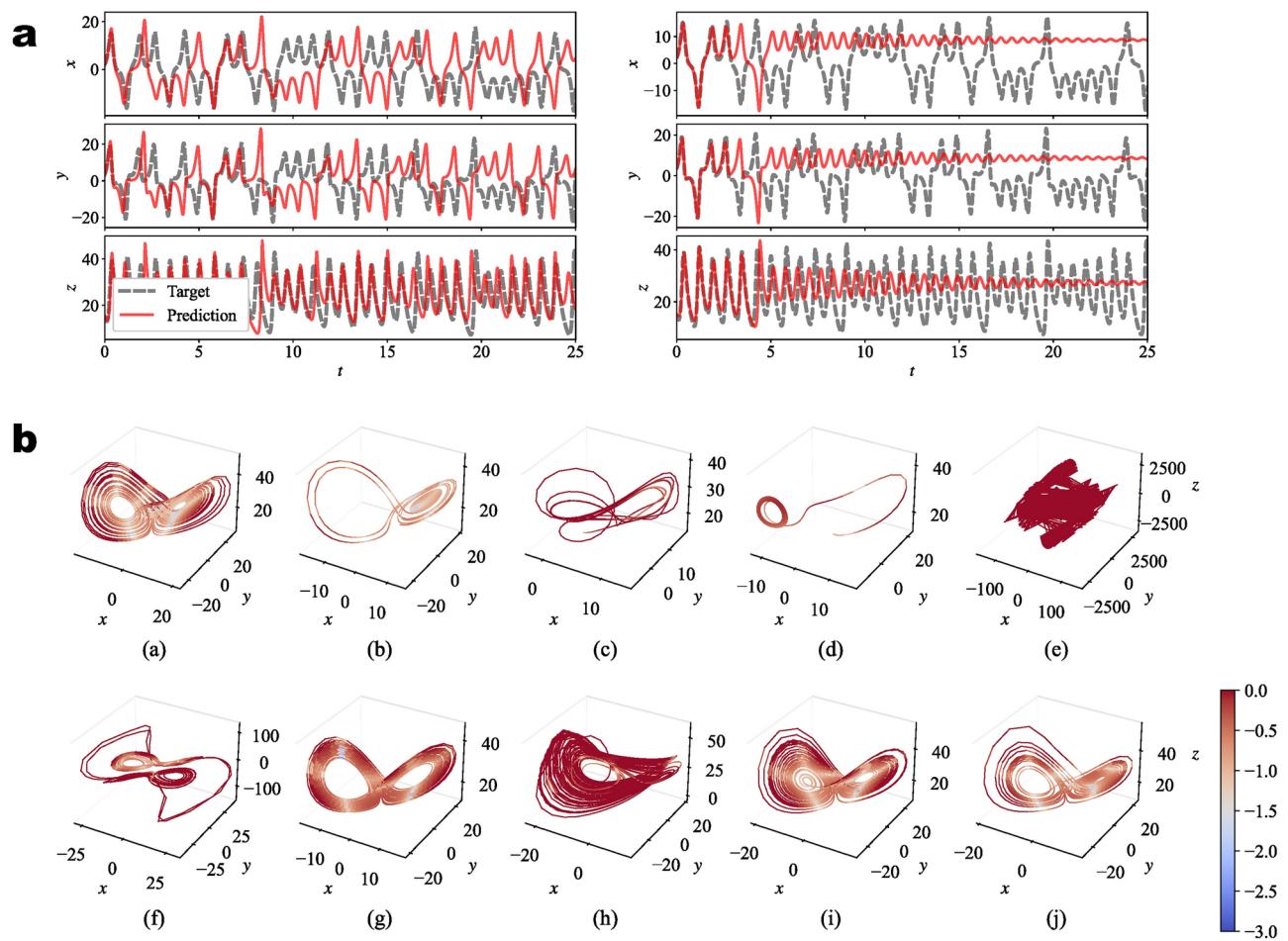


Fig. 4. Long-term prediction (closed-loop) using the autonomous \mathcal{L} -ESN. (a) The time series of the target (grey dashed) and prediction (red solid). The difference between the left and right panels lies in the realizations of the random numbers used for A and B . (b) The phase space structures of the orbits generated by the autonomous \mathcal{L} -ESN. The panels (a)–(j) are results corresponding to the ten times realizations of the random numbers used for A and B . The colors represent the local conjugacy error, \mathcal{E}_t^c . The length of the orbits shown is $T = 50$.

where the black dashed lines and the red solid lines represent the time series of the target Lorenz chaos and the prediction by the autonomous \mathcal{L} -ESN, respectively. For $t < 0$ we use the open-loop method, and switch to the closed-loop method at $t = 0$. For both results, the orbits by the automated \mathcal{L} -ESN deviate significantly from the target orbits at $t \gtrsim 3$, which is about three Lyapunov times, since the maximal Lyapunov exponent is $\lambda = 0.906$. The second case, shown in the right panel of Fig. 4a, suggests that the dynamics generated by the autonomous \mathcal{L} -ESN is not chaotic, but converges to a fixed point, which is unstable in the target Lorenz system.

To investigate the reconstruction ability, we demonstrate the phase space structure of the orbit generated by the autonomous \mathcal{L} -ESN for 10 different realizations in (a)–(j) of Fig. 4b, where the left and right panels of Fig. 4a correspond to the cases of (a) and (b), respectively. Obviously, the reconstruction ability of the automated \mathcal{L} -ESN is highly dependent on the realizations; in other words, it is not robust.

Figure 5 is the same as Fig. 4, but we use the \mathcal{Q} -ESN instead of the \mathcal{L} -ESN. The automated \mathcal{Q} -ESN exhibits the long-term prediction ability over about 8 Lyapunov times, which is remarkably improved compared to the case of the \mathcal{L} -ESN used, even with the same network size. Due to the intrinsic orbital instability of the Lorenz chaos, the orbits generated by the automated \mathcal{Q} -ESN inevitably deviate from the target orbits; however, the phase space structures shown in Fig. 5b are qualitatively equivalent to the Lorenz attractor. As will be quantitatively examined later, the reconstruction ability of the automated \mathcal{Q} -ESN is independent of the realizations; in other words, it can robustly reproduce the dynamics of Lorenz chaos. Similar results have been reported in the previous studies, e.g. Pathak et al.⁵; however, they used the relatively large network such as $N = 300$. We emphasize that the autonomous \mathcal{Q} -ESN has such a long-term prediction and robust reconstruction ability with the tiny network, $N = 10$.

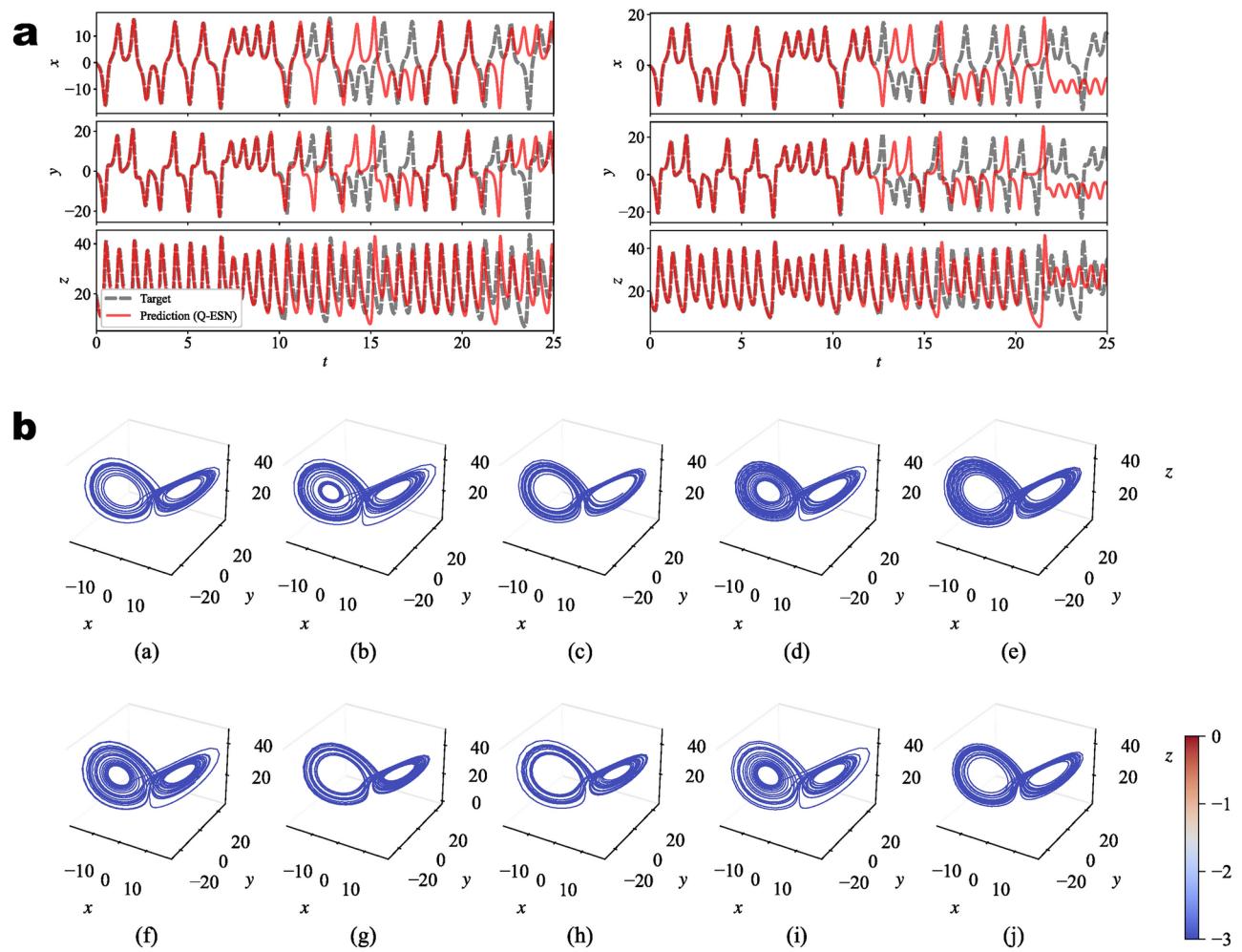


Fig. 5. Long-term prediction (closed-loop) using the autonomous \mathcal{Q} -ESN. The same as Fig. 4, but the \mathcal{Q} -ESN is used instead of the \mathcal{L} -ESN.

Quantitative comparison

For quantitative comparison, we introduce the mean conjugacy error (henceforth MCE), and the Kullback-Leibler divergence (henceforth KLD), which quantify the error between orbits and the error between invariant distributions, respectively. First, we define the MCE. As discussed in Hara and Kokubu⁷, the dynamical system determined by the autonomous ESN, (G, \mathbb{R}^N) , is expected to be smoothly conjugate to the Lorenz dynamics, (ϕ, \mathbb{R}^K) . The MCE quantifies the deviation from the expected conjugacy as follows. The above relationship $\hat{y}_t = \hat{h}(r_t) \simeq y_t = x_{t+1} = \phi(x_t)$ implies

$$\hat{h}(r_t) \simeq \phi(x_t) \simeq \phi(\hat{h}(r_{t-1})). \quad (16)$$

On the other hand, for the autonomous ESN, we have $r_t = G(r_{t-1})$, leading to

$$\hat{h}(r_t) = \hat{h}(G(r_{t-1})). \quad (17)$$

Considering the map \hat{h} as the conjugacy map, we define the conjugacy error at the reservoir state $r_t \in \mathbb{R}^N$ by

$$\mathcal{E}_t^c := \|\phi(\hat{h}(r_t)) - \hat{h}(G(r_t))\|, \quad (18)$$

where ϕ is approximated by the four-stage and fourth-order Runge-Kutta method. The long-time average of \mathcal{E}_t^c along the orbit $\{r_t\}$ generated by G defines the MCE,

$$\bar{\mathcal{E}}^c := \langle \mathcal{E}_t^c \rangle_T \quad (19)$$

The colors of the orbits in Figs. 4b and 5b represent the conjugacy error, where the values on the color bars correspond to $\log \mathcal{E}_t^c$. Obviously, compared to the autonomous \mathcal{L} -ESNs (Fig. 4b), the colors of the orbits

generated by the autonomous \mathcal{Q} -ESNs (Fig. 5b) are blue almost everywhere on the attractor, suggesting that successful conjugacy to the target Lorenz dynamics.

While the MCE quantifies the reconstruction ability of the autonomous ESN, it is not perfect. For instance, if the orbit generated by G converges to the saddle point along the stable manifold of the target system, the MCE may take a small value. However, the saddle point cannot be an attractor. Therefore, in this case, the autonomous ESN fails to reproduce the target attractor, even if the MCE is small. To shed light on the ergodic aspect of the dynamics, we compare the invariant probability measures through D_{KL} as a quantification complementary to MCE.

Figure 6 shows the probability density functions (PDF) of the variable x . The grey dashed and red solid lines represent the PDF $p(x)$ calculated from the target Lorenz chaos data and the PDF $q(x)$ calculated from the autonomous ESN data, respectively. The two panels of Fig. 6a show the results of the autonomous \mathcal{L} -ESN, corresponding to the two cases shown in Fig. 4a. Although the time series shown in the left panel of Fig. 4a and the phase space structure shown in Fig. 4b (a) are similar to the target Lorenz, the PDF $q(x)$ shown in the left panel of Fig. 6a differs from $p(x)$. The time series shown in the right panel of Fig. 4a converges to the fixed point, resulting in the PDF $q(x)$ shown in the right panel of Fig. 6a having a delta function-like form, and apparently differing from $p(x)$.

The two panels of Fig. 6b show the results of the automated \mathcal{Q} -ESN, corresponding to the two cases shown in Fig. 5a. The PDFs $p(x)$ and $q(x)$ are quite similar, suggesting that the autonomous \mathcal{Q} -ESN can reproduce the global structure of the target Lorenz attractor, in addition to the accurate prediction along the orbit verified by the MCE, which is local in phase space. The values of D_{KL} , where its definition is $D_{KL}(p\|q) := - \int p(x) \ln(q(x)/p(x))dx$, are $D_{KL}(p\|q) = 1.7 \times 10^{-6}$ and 5.2×10^{-6} for the PDFs $q(x)$ shown in the left and right panels of Fig. 6b, respectively. These values of D_{KL} are significantly smaller than the values of D_{KL} in the case of the \mathcal{L} -ESN used, e.g. $D_{KL}(p\|q) = 4.8 \times 10^{-2}$ for the PDFs $q(x)$ shown in the left panel of Fig. 6a.

Figure 7 summarizes the quantitative comparison. The top and bottom panels show the MCE $\bar{\mathcal{E}}^c$ and KLD D_{KL} values over ten times realizations of the random matrices A and B , respectively. The left points, labelled “1” on the horizontal axis, represent the results for the autonomous \mathcal{L} -ESN, excluding the two extremely poor results shown in Fig. 4b (e) and (f). The centre points, labelled “2” on the horizontal axis, represent the results for the autonomous \mathcal{Q} -ESN, illustrating that the values of both the MCE $\bar{\mathcal{E}}^c$ and the KLD D_{KL} are significantly smaller

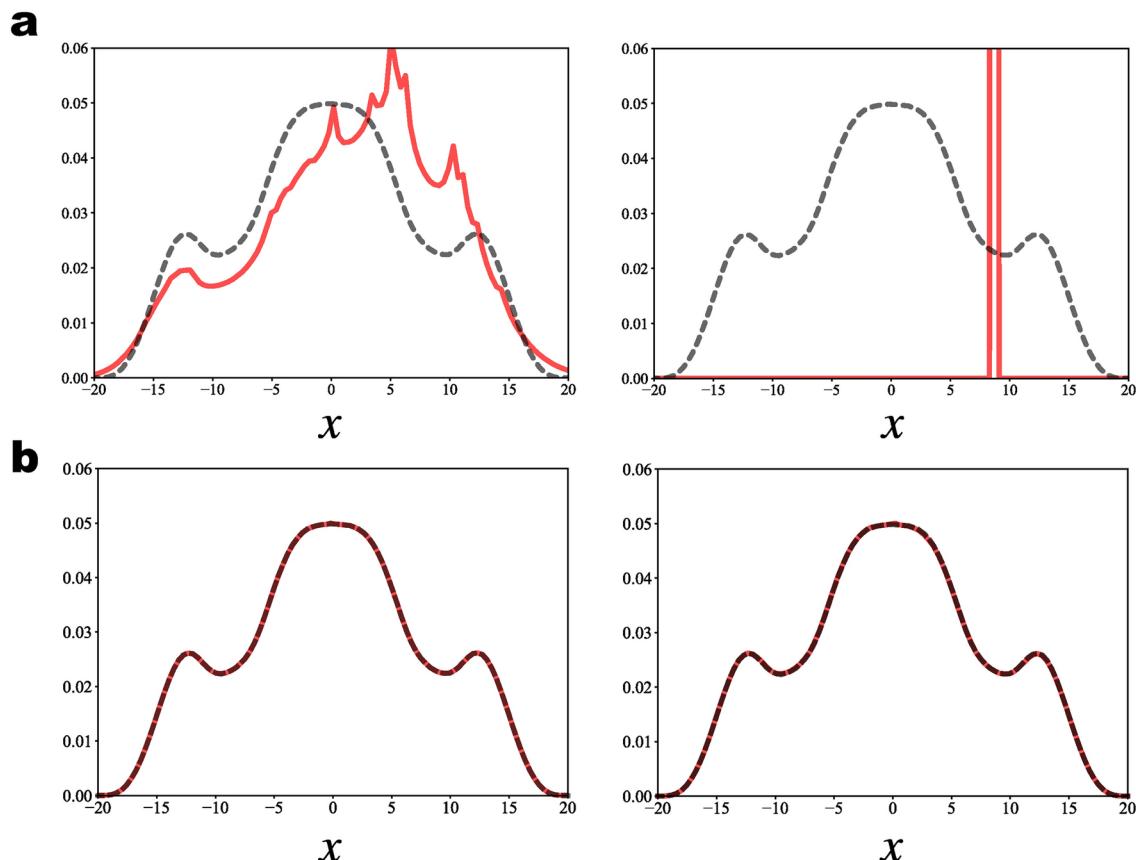


Fig. 6. Probability density functions (PDF) of the variable x . The dashed lines show the PDF of the target Lorenz system $p(x)$. The red solid lines show the PDF $q(x)$ calculated from data generated by (a), \mathcal{L} -ESN and (b), \mathcal{Q} -ESN. Two panels of (a) and (b) correspond to the cases shown in Figs. 4a and 5a, respectively.

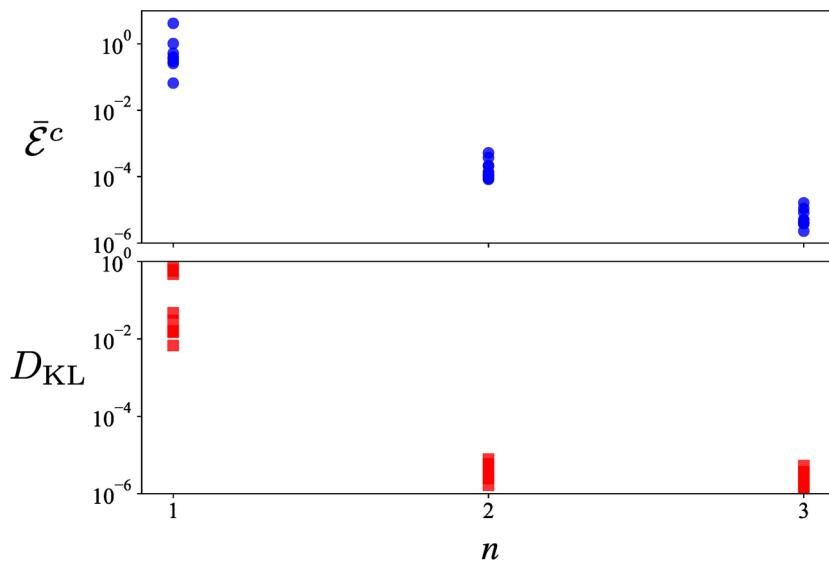


Fig. 7. Summary of quantitative comparison of reconstruction ability. The left ($n = 1$), center ($n = 2$), and right ($n = 3$) points correspond to the results of \mathcal{L} -, \mathcal{Q} -, and \mathcal{C} -ESN, respectively. The top and bottom panels show the MCE $\bar{\mathcal{E}}^c$ and KLD D_{KL} values over ten times realizations of the random matrices A and B , respectively.

than those of the \mathcal{L} -ESN. Immediately we notice this remarkable reconstruction ability of the \mathcal{Q} -ESN. Moreover, we find that both quantities are less dependent on the realizations of the random matrices compared to the \mathcal{L} -ESN, i.e. the \mathcal{Q} -ESN improves not only the accuracy but also the robustness of the reconstruction results.

Finally, we remark the results beyond the \mathcal{Q} -ESN, i.e. the cubic-form ESN (\mathcal{C} -ESN) including up to the third order terms of the reservoir variables, $r_j r_k r_\ell$, and the corresponding output weights, $W_{ijk\ell}^C$, which are trained to approximate the fourth term in the Taylor expansions (10) as

$$W_{ijk\ell}^C \simeq \frac{1}{6} \frac{\partial^3 h_i(0)}{\partial r_j \partial r_k \partial r_\ell}. \quad (20)$$

We show the MCE $\bar{\mathcal{E}}^c$ and the KLD D_{KL} for the autonomous \mathcal{C} -ESN in the right points, labelled “3” on the horizontal axis, of Fig. 7. As expected, the accuracy of the reconstruction by the autonomous \mathcal{C} -ESN is superior to that of the \mathcal{L} -ESN and the \mathcal{Q} -ESN. Furthermore, we find again that the \mathcal{C} -ESN improves not only the accuracy but also the robustness of the reconstruction results, compared to those of the \mathcal{L} -ESN and the \mathcal{Q} -ESN.

Conclusion and discussion

Inspired by the seminal works on the mathematical analysis of RC^{7,17}, we have proposed a novel method of RC with generalized readout with a theoretical guarantee of its high computational capabilities based on generalized synchronization. Numerical studies on the Lorenz and Rössler chaos have uncovered significant short- (Figs. 2 and 3) and long-term prediction and reconstruction abilities with improved robustness (Figs. 4, 5, 6 and 7) of the \mathcal{Q} -ESN. The MCE $\bar{\mathcal{E}}^c$ and KLD D_{KL} have quantified these properties complementarily, i.e. from the notions of orbit and distribution. By including the higher-order approximation, we have revealed “hierarchical” improvement in reconstruction ability and robustness; i.e. the \mathcal{C} -ESN is superior to the \mathcal{Q} -ESN, which is superior to the \mathcal{L} -ESN (Fig. 7).

As the future extensions based on the present work, we discuss the following three directions: mathematical analysis, machine learning, and physical implementation. From the mathematical analysis of RC^{7,17}, it may be natural that introducing the generalized readout improves prediction ability. However, we unexpectedly observed an improvement in the robustness of the reconstruction ability. Further analysis of the reservoir dynamics is crucial; unveiling fundamental properties such as the topological conjugacy and the mechanism behind the enhanced robustness will have major implications for several fields, including machine learning, where stabilizing the dynamics of neural networks by adding noise and normalization is one of the critical issues²⁰.

One of the key applications of the generalized readout is the physical RC; in many physical systems, such as photonic integrated circuit¹⁰, only small physical degrees of freedom are available^{9–13}. The hard challenge is to find a way to exploit these low-dimensional dynamics for computation. We emphasize that our generalized readout paves the way, and this is why we focus on small networks in the numerical study. For future work along the line of the research^{21,22}, combining linear physical systems with the generalized readout may be effective.

The apparent drawback of using the generalized readout is the large number of parameters to be trained, still within the linear learning framework. Therefore, based on the hierarchical improvement in accuracy with increasing parameters (Fig. 7), the balance between accuracy and learning cost should be determined for each application. For the large number of parameters to be trained, transfer learning^{23,24} may be efficient. Once linear regression is used, the trained parameters, i.e. the generalized readout weights, can be reused with a minor correction for similar tasks, e.g. predicting chaotic dynamics that are structurally stable.

The autonomous RC with generalized readout achieves accurate predictions, e.g. longer than 8 Lyapunov times (Fig. 4); however, such a prediction eventually fails due to the orbital instability. Toward practical predictions of, for instance, fluid turbulence²⁵, combining the autonomous RC with generalized readout and data assimilation may be essential in future work. Also, we have assumed that observational data of all state variables are available for training. It is important to investigate the effectiveness of the generalized readout in the case where only the partial observation data are available, as studied in²⁶. In practice, the prediction of high-dimensional dynamical systems, such as fluid turbulence^{23,27}, is crucial. We have found that the RC with generalized readout is effective for some high-dimensional chaos, which will be reported elsewhere.

The concept of generalized readout does not require the RC framework, but rather, may be essential in a more general machine learning context, e.g. training recurrent neural networks. Although the main claim of this paper is to propose the mathematical framework and the generalized readout method, a systematic comparison across a variety of neural networks, e.g., with deep architectures²⁸, in a more general task may be valuable and is left for future study. Studies of neural connections similar to Q - and C -ESN may also be interesting in the context of a learning mechanism in biological brains. Whatever the direction, the concepts from dynamical system theory used in the above discussion, such as synchronization, orbital instability, and conjugacy, will shed light on a guiding principle for future studies.

Data availability

The program used to generate the data is provided within the supplementary information file.

Received: 17 June 2024; Accepted: 29 November 2024

Published online: 28 December 2024

References

- Jaeger, H. *The “echo state” approach to analysing and training recurrent neural networks-with an erratum note*. Bonn, Germany: German National Research Center For Information Technology GMD Technical Report vol. 148, 13 (2001)
- Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560 (2002).
- Nakajima, K. & Fischer, I. *Reservoir Computing: Theory, Physical Implementations, and Applications, Natural Computing Series* (Springer, Berlin, 2021).
- Jaeger, H. & Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80 (2004).
- Pathak, J., Lu, Z., Hunt, B., Girvan, M. & Ott, E. Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data. *Chaos Interdiscip. J. Nonlinear Sci.* **27**(9), 121102 (2017)
- Kim, J., Lu, Z., Nozari, E., Pappas, G. & Bassett, D. Teaching recurrent neural networks to infer global temporal structure from local examples. *Nat. Mach. Intell.* **3**, 316–323 (2021).
- Hara, M. & Kokubu, H. Learning dynamics by reservoir computing. *J. Dyn. Differ. Equ.* **36**, 515–540 (2022).
- Kobayashi, M., Nakai, K., Saiki, Y. & Tsutsumi, N. Dynamical system analysis of a data-driven model constructed by reservoir computing. *Phys. Rev. E* **104**, 044215 (2021).
- Wang, S. et al. Others Echo state graph neural networks with analogue random resistive memory arrays. *Nat. Mach. Intell.* **5**, 104–113 (2023).
- Takano, K. et al. Compact reservoir computing with a photonic integrated circuit. *Opt. Express* **26**, 29424–29439 (2018).
- Sunada, S. & Uchida, A. Photonic reservoir computing based on nonlinear wave dynamics at microscale. *Sci. Rep.* **9**, 19078 (2019).
- Appeltant, L. et al. Information processing using a single dynamical node as complex system. *Nat. Commun.* **2**, 468 (2011).
- Tanaka, G. et al. Recent advances in physical reservoir computing: A review. *Neural Netw.* **115**, 100–123 (2019).
- Sande, G., Brunner, D. & Soriano, M. Advances in photonic reservoir computing. *Nanophotonics* **6**, 561–576 (2017).
- Inubushi, M. & Yoshimura, K. Reservoir computing beyond memory–nonlinearity trade-off. *Sci. Rep.* **7**, 10199 (2017).
- Inubushi, M., Yoshimura, K., Ikeda, Y. & Nagasawa, Y. On the characteristics and structures of dynamical systems suitable for reservoir computing. *Reserv. Comput. Theory Phys. Implement. Appl.* 97–116 (2021).
- Grigoryeva, L., Hart, A. & Ortega, J. Chaos on compact manifolds: Differentiable synchronizations beyond the Takens theorem. *Phys. Rev. E* **103**, 062204 (2021).
- Hertaux, J. & Rath, C. Breaking symmetries of the reservoir equations in echo state networks. *Chaos: Interdiscip. J. Nonlinear Sci.* **30**(13), 123142 (2020).
- Boltt, E. On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD. *Chaos Interdiscip. J. Nonlinear Sci.* **31**(23), 013108 (2021).
- Wikner, A. et al. Stabilizing machine learning prediction of dynamics: Novel noise-inspired regularization tested with reservoir computing. *Neural Netw.* **170**, 94–110 (2024).
- Shougat, M., Li, X., Mollik, T. & Perkins, E. An information theoretic study of a duffing oscillator array reservoir computer. *J. Comput. Nonlinear Dyn.* **16**, 081004 (2021).
- Coulombe, J., York, M. & Sylvestre, J. Computing with networks of nonlinear mechanical oscillators. *PLoS ONE* **12**, e0178663 (2017).
- Inubushi, M. & Goto, S. Transfer learning for nonlinear dynamics and its application to fluid turbulence. *Phys. Rev. E* **102**, 043301 (2020).
- Sakamaki, R., Kanno, K., Inubushi, M. & Uchida, A. Transfer learning based on photonic reservoir computing using semiconductor laser with optical feedback. *IEICE Proc. Ser.* **71**, 229–232 (2022).
- Inubushi, M., Saiki, Y., Kobayashi, M. & Goto, S. Characterizing small-scale dynamics of Navier–Stokes turbulence with transverse Lyapunov exponents: A data assimilation approach. *Phys. Rev. Lett.* **131**, 254001 (2023).
- Storm, L., Gustavsson, K. & Mehlig, B. Constraints on parameter choices for successful time-series prediction with echo-state networks. *Mach. Learn. Sci. Technol.* **3**, 045021 (2022).

27. Matsumoto, S., Inubushi, M. & Goto, S. Stable reproducibility of turbulence dynamics by machine learning. *Phys. Rev. Fluids* **9**, 104601 (2024).
28. Wang, R., Kalnay, E. & Balachandran, B. Neural machine-based forecasting of chaotic dynamics. *Nonlinear Dyn.* **98**, 2903–2917 (2019).

Acknowledgements

We thank M. Hara, H. Kokubu, S. Sunada, T. Yoneda, S. Matsumoto, S. Goto, Y. Saiki, and J. A. Yorke for their insightful comments and encouragement. We would also like to thank C. P. Caulfield and DAMTP, University of Cambridge, for providing a great environment in which this work was completed on M.I.'s sabbatical. This work was partially supported by JSPS Grants-in-Aid for Scientific Research (Grants Nos. 22K03420, 22H05198, 20H02068 and 19KK0067).

Author contributions

M.I. and A.O. wrote the main manuscript text and prepared all the figures. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-81880-3>.

Correspondence and requests for materials should be addressed to M.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024