Model-Size Reduction for Reservoir Computing by Concatenating Internal States Through Time

Yusuke Sakemi (sakemi@iis.u-tokyo.ac.jp)
The University of Tokyo
Kai Morino
Kyushu University,
Timothee Leleu
The University of Tokyo
Kazuyuki Aihara
The University of Tokyo

Article

Keywords: Reservoir computing, implementations of RC, reduce reservoir size, edge computing

Posted Date: July 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-40208/v1

Model-Size Reduction for Reservoir Computing by Concatenating Internal States Through Time

Yusuke Sakemi * [a,b], Kai Morino[a,c], Timothée Leleu[a,d], and Kazuyuki Aihara[a,d]

[a]Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba Meguro-ku, Tokyo 153-8505, Japan

[b]NEC Corporation, 1753 Shimonumabe Nakahara-ku, Kanagawa 211-8666, Japan

[c]Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, 6-1 KasugaKouen, Kasuga-shi, Fukuoka 816-8580, Japan

[d]International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo 113-0033, Japan

July 6, 2020

Abstract

Reservoir computing (RC) is a machine learning algorithm that can learn complex time series from data very rapidly based on the use of high-dimensional dynamical systems, such as random networks of neurons, called "reservoirs." To implement RC in edge computing, it is highly important to reduce the amount of computational resources that RC requires. In this study, we propose methods that reduce the size of the reservoir by inputting the past or drifting states of the reservoir to the output layer at the current time step. To elucidate the mechanism of model-size reduction, the proposed methods are analyzed based on information processing capacity proposed by Dambre *et al.* (2012). In addition, we evaluate the effectiveness of the proposed methods on time-series prediction tasks: the generalized Hénon-map and NARMA. On these tasks, we found that the proposed methods were able to reduce the size of the reservoir up to one tenth without a substantial increase in regression error.

Efficiently processing time-series data is important for various tasks, such as time-series forecasting, anomaly detection, natural language processing, and system control. Recently, machine-learning approaches for these tasks have attracted much attention of researchers and engineers because they not only require little domain knowledge but also often perform better than traditional approaches. In particular, machine-learning models that employ recurrent neural networks, such as long short-term memory, have achieved great success in natural language processing and speech recognition [1], and their fields of applications continue to expand. However, the standard learning algorithms for recurrent neural networks, which include backpropagation through time [2] and its variants [3], require large computational resources. These computational burdens often hinder real-world applications, especially when computing is performed near end users. Such computing is called "edge computing," and characterized by limited computational power and limited battery capacity.

Reservoir computing (RC) is a machine-learning algorithm that aims to reduce the computational resources required for predicting time series without reducing accuracy. As shown in Fig. 1, a typical RC consists of three parts: an input layer, a "*reservoir*" layer where neurons are randomly connected, and an output layer [4, 5]. Because only the weights between the reservoir layer and the output layer are trained while the other weights remain fixed, the learning process of RC is much faster than that of backpropagation through time [6, 7, 8]. Therefore, RC is expected to be a lightweight machine-learning algorithm that enables machine learning in edge computing [9].

The RC training is fast and accurate. In addition, RC has shown high performance on various time-series forecasting tasks; examples include chaotic time-series [10, 11, 12], weather [13], wind-power generation [14],

*sakemi@iis.u-tokyo.ac.jp

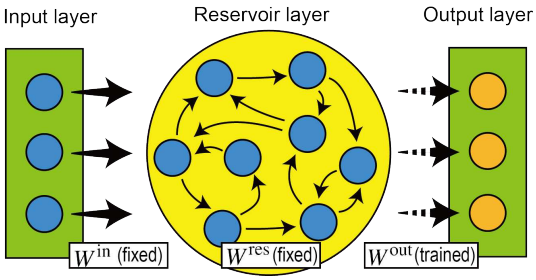

Figure 1: **Typical RC architecture.** The reservoir layer consists of randomly connected neurons. The connections between the input and reservoir layer W^{in} and connections within the reservoir layer W^{res} are fixed (solid arrows), whereas the output weights W^{out} are trained (dashed arrows).

and finance [15]. Moreover, the range of applications of RC has extended into control engineering [16, 17] and video processing [18, 19, 20].

To develop the applications of RC in edge computing, its hardware implementation must be improved to enhance its computational speed and energy efficiency. For realizing such efficient hardware implementation, variants of RC models, some of which employ delay-feedback systems [21], simple network topologies such as ring-topology and delay lines [22, 23, 24], and billiard systems [25], have been proposed. Efficient hardware based on these variants have been implemented using field programmable integrated circuits (FPGAs) [26, 27, 28, 29]. Moreover, numerous types of implementation employing physical systems, such as photonics [30, 31, 32], spintronics [33], mechanical oscillators [34], and analog integrated electronic circuits [35, 36], have been demonstrated [37]. Although these implementations have exhibited the superiority of RC in computational speed and energy efficiency, the maximum size of the reservoir and, in turn, the forecasting accuracy, is limited by the physical size of the hardware.

In this study, we propose three methods that reduce the size of the reservoir without any performance impairment. The three methods share the concept that the number of the effective dimension of the reservoir is increased by allowing additional connections from the reservoir layer at multiple time steps to the output layer at the current time step. We analyze the mechanism of the proposed methods based on the *information processing capacity* (IPC) proposed by Dambre *et al.* [38]. We also demonstrate how the proposed methods reduce the size of the reservoir in the generalized Hénon-map and NARMA tasks.

Results

RC framework.

In the mathematical representation of RC, four vector variables are defined as follows: $\boldsymbol{u}(t) \in \mathbb{R}^{N^{\text{in}}}$ for the inputs, $\boldsymbol{x}(t) \in \mathbb{R}^{N^{\text{res}}}$ for the states of the reservoir, $\boldsymbol{y}(t) \in \mathbb{R}^{N^{\text{out}}}$ for the outputs, and $\boldsymbol{y}^{\text{tc}}(t) \in \mathbb{R}^{N^{\text{out}}}$ for the teaching signals. The constants $N^{\text{in}}, N^{\text{res}}$, and N^{out} are the dimensions of the inputs, states of the reservoir, and outputs, respectively. The updates of the reservoir states are given by

$$\boldsymbol{x}(t) = \tanh\left(W^{\text{res}}\boldsymbol{x}(t-1) + W^{\text{in}}\boldsymbol{u}(t)\right), \tag{1}$$

where $W^{\text{in}} \in \mathbb{R}^{N^{\text{res}} \times N^{\text{in}}}$ is a weight matrix representing the connections from the neurons in the input layer to those in the reservoir layer. Its elements are independently drawn from uniform distribution $U(-\rho^{\text{in}}, \rho^{\text{in}})$, where ρ^{in} is a positive constant. Another weight matrix $W^{\text{res}} \in \mathbb{R}^{N^{\text{res}} \times N^{\text{res}}}$ represents the connections among the neurons in the reservoir layer. Its elements are initialized by drawing values from uniform distribution $U(-1, 1)$ and subsequently divided by a positive value to ensure that the spectral radius of W^{res} is ρ^{res}. Note that elements in matrices W^{in} and W^{res} are fixed to the initialized values. The outputs are obtained by

$$\boldsymbol{y}(t) = W^{\text{out}}\boldsymbol{x}(t), \tag{2}$$

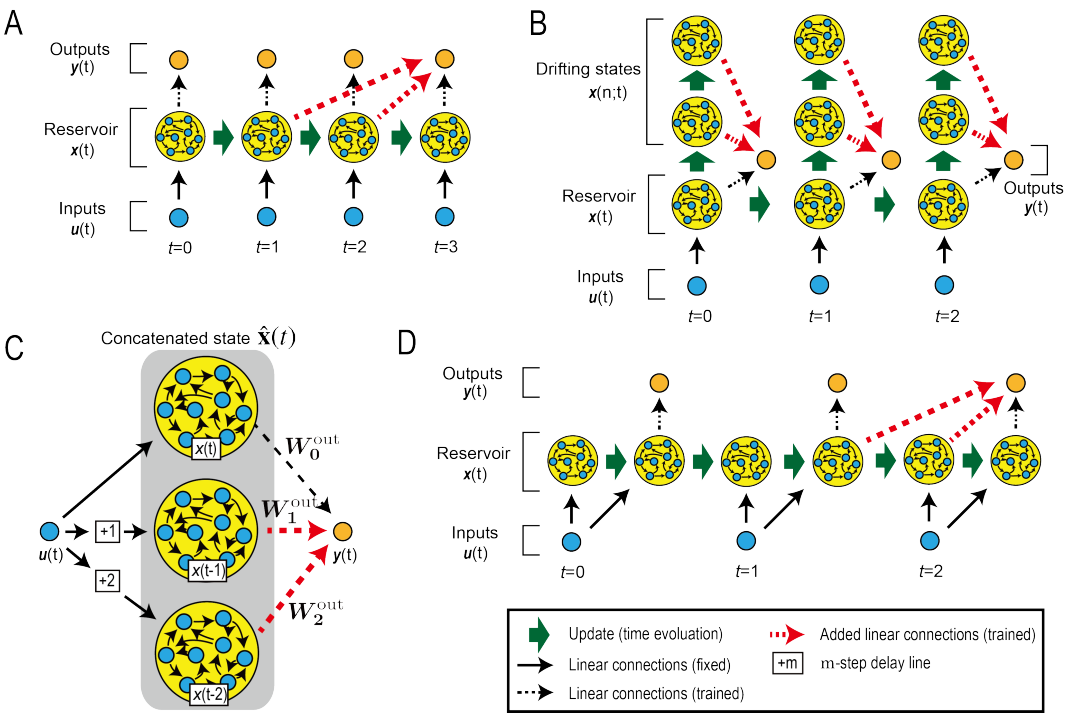

Figure 2: **Schematics of the proposed methods.** A. Delay-state concatenation when the number of additional connections P is two and the unit of delay Q is one. B. Drift-state concatenation when P is two and Q is one. C. Another view of delay-state concatenation. The reservoir consists of three identical dynamical systems and delay lines. The added dynamical systems have +1 delay lines and +2 delay lines, respectively. D. Delay-state concatenation with one transient state when P is two and Q is one.

where $W^{\text{out}} \in \mathbb{R}^{N^{\text{out}} \times N^{\text{res}}}$ is a weight matrix representing connections from the neurons in the reservoir layer to those in the output layer. The output weight matrix W^{out} is trained in the offline learning process of RC by using the pseudoinverse (see Methods sections).

Proposed methods.

We propose three methods that modify the connections between the reservoir and output layers. We call these three methods (i) *delay-state concatenation*, (ii) *drift-state concatenation*, and (iii) *delay-state concatenation with transient states*. These methods share the idea that the number of the effective dimension of the reservoir is increased by allowing additional connections from the reservoir layer at multiple time steps to the output layer at the current time step. For the delay-state concatenation and delay-state concatenation with transient states, additional connections are formed from the past states of the reservoir layer to the current output layer, as illustrated in Figs. 2 A and D. On the other hand, for the drift-state concatenation, additional connections are formed from newly introduced states of the reservoir, called drifting states, to the current output layer, as illustrated in Fig. 2 B. The drifting states are obtained by updating the current states of the reservoir layer without input signals. In what follows, we mathematically formulate these three proposed methods.

First, we formulate the delay-state concatenation with concatenated states of the reservoir given by

$$\hat{\boldsymbol{x}}(t) := \begin{pmatrix} \boldsymbol{x}(t) \\ \boldsymbol{x}(t-Q) \\ \vdots \\ \boldsymbol{x}(t-PQ) \end{pmatrix}. \tag{3}$$

Note that $\boldsymbol{x}(t)$ is a column vector and the number of neurons in the reservoir does not change. A positive integer Q represents the unit of delays. Another positive integer P represents the number of past states that are concatenated to the current states. The outputs are obtained with the concatenated states $\hat{\boldsymbol{x}}(t)$ and the corresponding output-weight matrix $\hat{W}^{\text{out}}$ as follows:

$$\hat{W}^{\text{out}} := \left(W_0^{\text{out}}\ W_1^{\text{out}}\ \cdots\ W_P^{\text{out}}\right), \tag{4}$$

$$\begin{aligned}\boldsymbol{y}(t) &= \sum_{i=0}^{P} W_i^{\text{out}} \boldsymbol{x}(t - iQ) \\ &= \hat{W}^{\text{out}} \hat{\boldsymbol{x}}(t).\end{aligned} \tag{5}$$

Here, $\hat{\boldsymbol{x}}(t)$ and $\hat{W}^{\text{out}}$ are defined in $\mathbb{R}^{(P+1)N^{\text{res}}}$ and $\mathbb{R}^{N^{\text{out}} \times (P+1)N^{\text{res}}}$, respectively. Figure 2 A shows a schematic of this method illustrating the prediction of $\boldsymbol{y}(3)$ when $Q = 1$ and $P = 2$. One can see that there are additional connections from the past states of the reservoirs $\boldsymbol{x}(1)$ and $\boldsymbol{x}(2)$ to output $\boldsymbol{y}(3)$, as indicated by red dashed arrows. From a different point of view, this model can be illustrated using the concatenated states of the reservoir $\hat{\boldsymbol{x}}(t)$, as in Fig. 2 C, where the reservoir consists of three identical smaller reservoirs, each with a different time delay of 0, 1, and 2 from the inputs. Evidently, the effective dimension of the concatenated reservoir is three times larger than that of the original reservoir. Because the learning performance is enhanced by using a larger reservoir [7], the proposed method should be able to increase the computing capability without needing to add neurons in the reservoir.

Second, we formulate drift-state concatenation, as illustrated in Fig. 2 B by introducing the drifting states of the reservoir given by

$$\boldsymbol{x}^{\text{drift}}(t'; t) = \begin{cases} \tanh\left(W^{\text{drift}} \boldsymbol{x}(t)\right), & (\text{if } t' = 1), \\ \tanh\left(W^{\text{drift}} \boldsymbol{x}^{\text{drift}}(t' - 1; t)\right), & (\text{if } t' \geq 2), \end{cases} \tag{6}$$

where $\boldsymbol{x}^{\text{drift}}(t'; t)$ represents the drifting states of the reservoir and t' is the time step after the current time step t. Using the drifting states, we redefine the concatenated states of the reservoir and the corresponding output matrix as follows:

$$\hat{\boldsymbol{x}}(t) := \begin{pmatrix} \boldsymbol{x}(t) \\ \boldsymbol{x}^{\text{drift}}(1; t) \\ \boldsymbol{x}^{\text{drift}}(2; t) \\ \vdots \\ \boldsymbol{x}^{\text{drift}}(P; t) \end{pmatrix}, \tag{7}$$

$$\hat{W}^{\text{out}} := \left(W_0^{\text{out}}\ W_1^{\text{out}}\ \cdots W_P^{\text{out}}\right), \tag{8}$$

$$\begin{aligned}\boldsymbol{y}(t) &= W_0^{\text{out}} \boldsymbol{x}(t) + \sum_{i=1}^{P} W_i^{\text{out}} \boldsymbol{x}^{\text{drift}}(i; t) \\ &= \hat{W}^{\text{out}} \hat{\boldsymbol{x}}(t).\end{aligned} \tag{9}$$

Here, W^{drift} is a matrix representing the connections within the reservoir to obtain the drifting states, and its elements are drawn from uniform distribution $U(-1, 1)$ divided by a positive value to ensure that the spectral radius of W^{drift} is ρ^{drift}.

Third, delay-state concatenation with transient states introduces transient states to the delay-state concatenation, as illustrated in Fig. 2 D where the states of the reservoir update twice (so it has one transient state) during the inputs and the outputs update once (see Methods section).

Although we have shown that the proposed methods can increase the effective dimension of the reservoir without adding neurons, one potential drawback of the methods is the cost of the memory required to store the past reservoir states. Delay-state concatenation is very memory efficient. To carry out RC with delay-state concatenation, the terms

$$W_0^{\text{out}} \boldsymbol{x}(t),\ W_1^{\text{out}} \boldsymbol{x}(t), \ldots, \text{ and } W_P^{\text{out}} \boldsymbol{x}(t) \tag{10}$$

are computed and stored in memory at time step t. The dimensions of these vectors are all N^{out}. The vector $W_i^{\mathrm{out}}\boldsymbol{x}(t)$ must be stored until they are used for calculating the outputs at time steps $t+iQ$. Therefore, the total memory cost is obtained as follows:

$$\begin{aligned} &N^{\mathrm{out}} + (Q+1)N^{\mathrm{out}} + (2Q+1)N^{\mathrm{out}} + \cdots + (PQ+1)N^{\mathrm{out}} \\ &= \frac{(P+1)(PQ+2)N^{\mathrm{out}}}{2}. \end{aligned} \tag{11}$$

It should be noted that this method increases the effective dimension of the reservoir state by a factor of $(P+1)$, but the number of the neurons in the reservoir is not increased. If the number of neurons is increased by $(P+1)$ times, the additional memory cost to store the reservoir states becomes PN^{res}. For moderate numbers of P and Q (typically less than 5), the memory cost for the proposed method given by Eq. 11 is much less than PN^{res} because $N^{\mathrm{out}} \ll N^{\mathrm{res}}$ for typical RC applications. The memory required to store the weights within the reservoir is also reduced from $(PN^{\mathrm{res}})^2 + PN^{\mathrm{out}}$ to $(N^{\mathrm{res}})^2 + PN^{\mathrm{res}}N^{\mathrm{out}}$ in the same way given that the reservoir is fully connected. Therefore, delay-state concatenation can increase the dimensions of the reservoir more efficiently than by simply increasing the number of neurons in the reservoir.

Quantitative analysis based on IPC.

Before benchmarking the proposed RC, we quantitatively analyze the learning capacity of the RC to elucidate how the proposed methods work.

The *memory capacity* (MC) is a performance measure commonly used in the RC research community [39]. The MC represents how precisely the RC system can reproduce the past inputs. A number of studies have shown that the MC is theoretically bounded by the number of neurons in the reservoir, and the MC can reach this bound in some situations [40, 23, 24]. Boedecker *et al.* [41] evaluated the MC at the edge of chaos, which is a region in the model parameter space where RC is stable but near to unstable. Farkaš *et al.* [42] have evaluated the MC for various model parameters. Ganguli *et al.* [43] extended the concept of MC using Fischer information. However, the MC does not evaluate how well RC processes information in a nonlinear way. Because many tasks in the real world targeted by RC are nonlinear problems, the MC is not a suitable measure for analyzing the proposed methods in this sense. Therefore, to elucidate the mechanism of the methods proposed in this study, we employed another criterion [38] called the information processing capacity (IPC), which handles nonlinear tasks.

The IPC is a measure that integrates both memory and information processing performance. By employing an orthogonal basis set that spans the Hilbert space, one IPC can be obtained from one corresponding basis. The IPC can be interpreted as a quantity that represents not only how well the network can memorize past inputs but also how precisely the network can convert inputs into the target outputs in a nonlinear manner given the basis set. Dambre *et al.* [38] showed that the total IPC C^{total}, which is a sum of all IPCs, is identical to the number of neurons in the reservoir, provided (I) the inputs are independent and identically distributed (*i.i.d.*), (II) the fading memory condition is satisfied, and (III) all the neurons are linearly independent (see Theorem 7 and its proof in [38]). By analyzing the IPCs, one can obtain a large amount of information about how RC processes input data. For example, the degree of nonlinearity of the information processing carried out in RC can be analyzed by calculating multi-order IPCs. The kth-order IPC $C^{k\mathrm{th}}$ is defined as the sum of the IPCs corresponding to the subset of a basis with kth-order nonlinearity. Based on the IPC, informative results such as the memory–nonlinearity tradeoff have been obtained [38]. Therefore, using the IPC, we can analyze how RC stores and processes information in the reservoir as well as how the proposed methods affect the way information is processed.

We calculated the IPCs for the standard RC and for RC with the proposed methods (see Methods section). Figure 3 shows the IPCs when $N^{\mathrm{res}} = 12$ and $N^{\mathrm{res}} = 24$ for various values of ρ^{in} and ρ^{res}. Note that only odd-order IPCs were observed because of the symmetry. In each setting, we calculated the IPCs for the standard RC (left columns), those for the RC with delay-state concatenation with $P = Q = 1$ (center columns), and those for the RC with drift-state concatenation with $P = Q = 1$ (right columns). One can find that the value of $C^{\mathrm{total}}/(P+1)$ almost reaches the number of neurons N^{res} in reservoir except when $\rho^{\mathrm{res}} = 1.05$. This result indicates that the proposed methods actually increase the total IPC by $(P+1)$ times. The observed lower values of the total IPCs for the case of $\rho^{\mathrm{res}} = 1.05$ can be attributed to the failure-of-fading-memory condition. In the RC research community, it is well known that the dynamics of

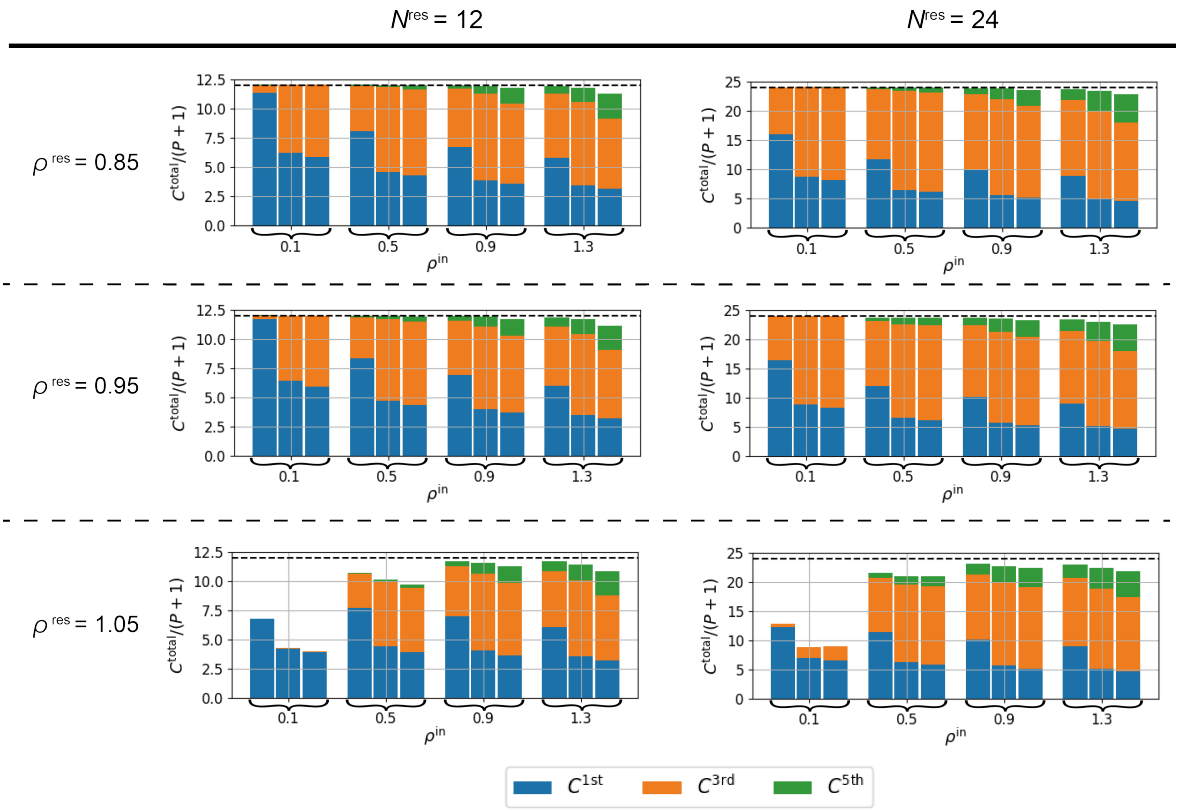

Figure 3: **Analysis of information processing capacities (IPCs).** IPCs of standard RC and RC with the proposed methods for various values of input weight strength ρ^{in} and spectral radius ρ^{res}. The left and right panels show the results for $N^{\text{res}} = 12$ and $N^{\text{res}} = 24$, respectively. The dashed horizontal line within each subgraph represents the value of N^{res}. For each value of ρ^{in}, the left column presents the standard RC with $P = 0$, the center column presents delay-state concatenation with $P = Q = 1$, and the right column presents drift-state concatenation with $P = Q = 1$.

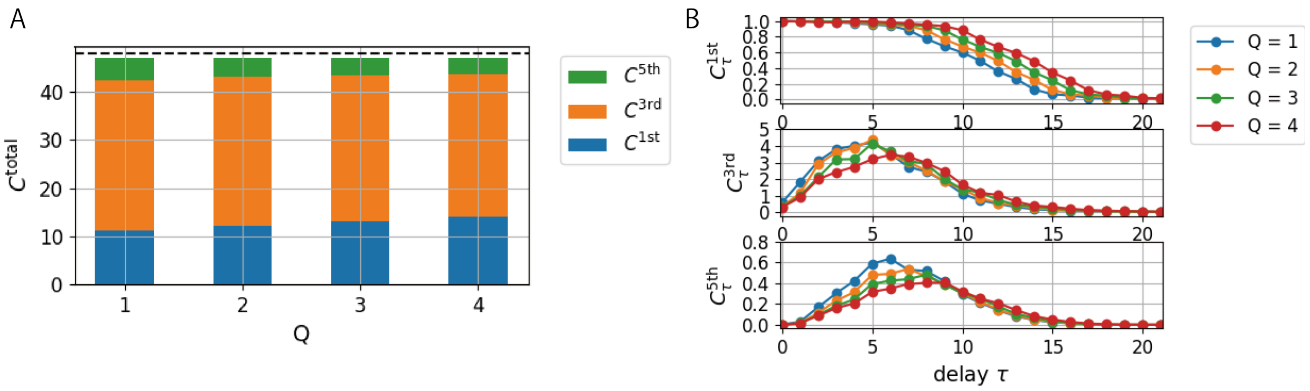

Figure 4: **Q-dependence of IPCs.** (A) IPCs of the RC with delay-state concatenation for various values of Q. (B) Delay structures of the IPCs for various values of Q. From top to bottom, the first-, third-, and fifth-order IPCs are shown. The parameters are set as follows: $N^{\text{res}} = 24$, $\rho^{\text{in}} = 0.9$, $\rho^{\text{res}} = 0.95$, and $P = 1$.

RC is more likely to be chaotic when ρ^{res} increases (typically occurring when ρ^{res} is larger than 1) [44], and this corresponds to the failure of fading memory. For all cases, as ρ^{in} increases, the third-order IPC $C^{3\text{rd}}$ and the fifth-order IPC $C^{5\text{th}}$ tend to increase, which reflects the increase in nonlinearity in the reservoir [38] given the selected basis set. Note that as ρ^{in} increases, the total IPC C_{total} tends to decrease, which is attributed to increases in the importance of higher-order IPCs (e.g., seventh-order and ninth-order IPCs).

In Fig. 4 (A), we show the IPCs for delay-state concatenation with $P = 1$ for several values of the unit of delay Q. As Q increases, the first-order IPC increases as well. This result may be trivial because RC with large Q can access the past states of the reservoir, rendering the reproduction of the past inputs easy. To investigate the effects of the unit of delay Q on IPCs, we decomposed the kth-order IPC $C^{k\text{th}}$ into components in terms of their delay such as $C_0^{k\text{th}}$, $C_1^{k\text{th}}$, and $C_2^{k\text{th}}$, which correspond to a different subset of the basis. Figure 4 (b) shows the distributions of the delay components for four values of Q under the same experimental conditions. As the values of Q increase, the distribution tends to shift to the right (larger delays) for each order of IPC. This fact indicates that, as demonstrated in the subsequent section, one can tune RC models by adjusting the value of Q according to the delay structure of the target tasks.

Next, for various values of P, we show the IPCs in Fig. 5 (A) for delay-state concatenation, drift-state concatenation, and delay-state concatenation with one transient state. For all three proposed methods, the contributions of higher-order IPCs tend to be dominant in the total IPCs as P increases. In Fig. 5 (B), we show the delay structures of the IPCs. Note that to clarify how the distributions of the delay structure change as the value of P changes, we used the normalized IPC $C_\tau^{n\text{th}}/(P+1)\sum_\tau C_\tau^{n\text{th}}$. The top panels in Fig. 5 (B) show that, as P increases, the distribution of the delay structure of the IPCs for delay-state concatenation tends to shift to the right (larger delays). Conversely, in the middle panels, the IPCs for drift-state concatenation do not change significantly. These results may be explained as follows: the increase in P for delay-state concatenation increases the memory of past inputs because of the additional connections from the past states of the reservoir, whereas the increase in P for drift-state concatenation does not increase the memory of the past inputs because drifting states are obtained from the current states of the reservoir. The bottom panels of this figure show that the distribution of IPCs for delay-state concatenation with one transient state tends to shift to larger delays as P increases. However, the delays in this distribution are smaller than the delays in the distribution obtained using delay-state concatenation. This difference may stem from the fact that the information of past states stored in the reservoir is more likely to be thrown away and to be replaced with that of more recent states in delay-state concatenation with one transient state because the RC model in this case carries out nonlinear transformation twice for each input (see Fig. 2 D).

Here, we present a short summary of the above experiments. We have numerically shown that the total IPCs divided by $P+1$ are almost independent of the values of Q and P, which is consistent with the theory in Ref. [38]. Furthermore, we have found that the importance among IPC components and the delay structure

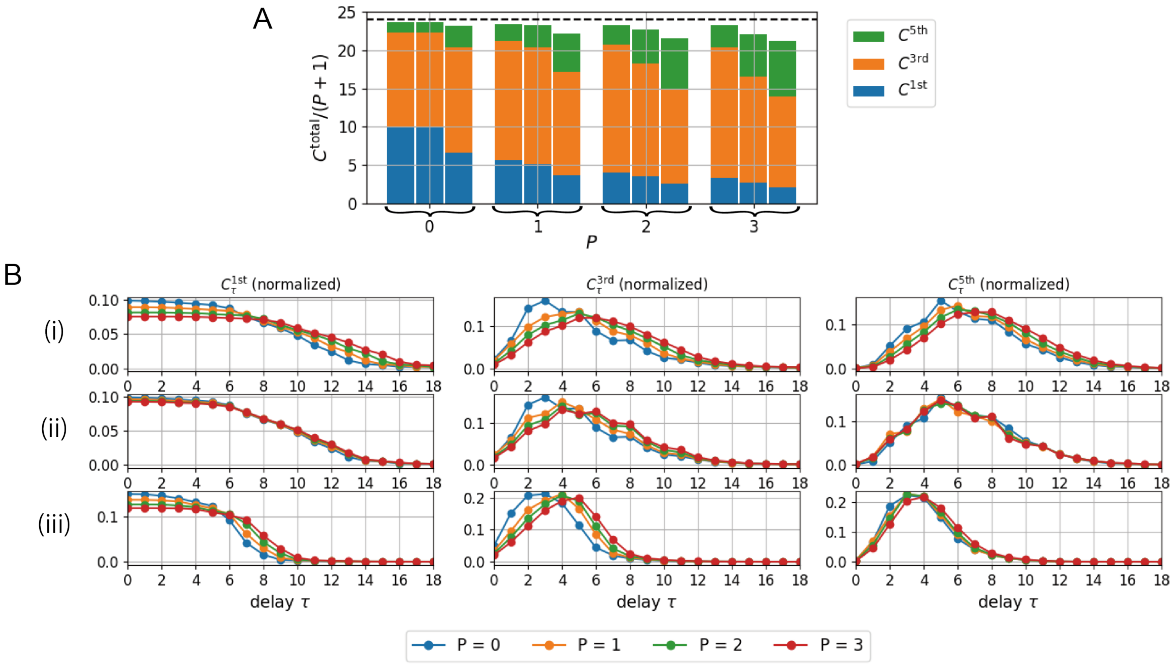

Figure 5: ***P*-dependence of IPCs.** (A) IPCs with the proposed methods for various values of P. Left columns: IPCs for delay-state concatenation, center columns: IPCs for drift-state concatenation, and right columns: IPCs for delay-state concatenation with one transient state. (B) Comparison of the delay structures of the IPCs for the three proposed methods. Top panels (i): delay structures for delay-state concatenation, middle panels (ii): delay structures for drift-state concatenation, and bottom panels (iii): delay structures for delay-state concatenation with one transient state. To highlight the difference in the distribution of the delay structures, the IPCs for each order are normalized. From left to right, the first-, third-, and fifth-order IPCs are shown. The parameters are set as follows: $N^{\mathrm{res}} = 24$, $\rho^{\mathrm{in}} = 0.9$, $\rho^{\mathrm{res}} = 0.95$, and $Q = 1$.

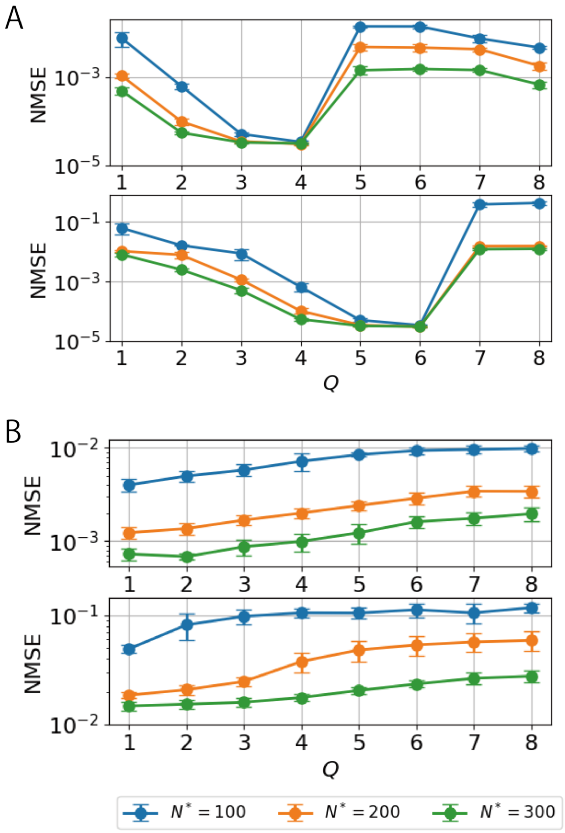

Figure 6: **Q-dependence of the learning performance.** (A) NMSE for a sixth-order Hénon-map task (the upper panel) and for an eighth-order Hénon-map task (the lower panel) for various values of Q with $P = 1$. (B) NMSE for NARMA5 (the upper panel) task and for NARMA10 (the lower panel) task for various values of Q with $P = 1$. Error bars show the standard deviation of the results of 10 trials.

of IPCs can be modified by selecting the values of Q and P. These findings indicate that the learning performance on real-world tasks may be enhanced by selecting appropriate values of Q and P adjusted to a target task with a specific temporal structure.

Effectiveness on Complex Data.

Although we have shown that the proposed methods can increase the IPCs efficiently, the conditions assumed above are not always guaranteed in real-world applications; for example, inputs may not be drawn from *i.i.d.* data, and neurons in the reservoir may not be linearly independent. Therefore, the IPCs are just a guide that help us understand the mechanisms of the proposed methods. In this section, to evaluate the effectiveness of the proposed methods on complex data, we applied them to two prediction tasks: generalized Hénon-map tasks and NARMA tasks (see Eq. 24 and Eq. 25 in Methods section). In the following experiments, the dimensions of the reservoir are approximately given by an integer N^*, and the actual number of neurons in the reservoir is given by $N^{\mathrm{res}} = \lfloor N^*/(P+1) \rfloor$, where $\lfloor m \rfloor := \max\{q \in \mathbb{Z} | q \leq m\}$. We optimized the model parameters, ρ^{in}, ρ^{res}, and ρ^{drift} with Bayesian optimization [45, 46].

We first investigated the effects of the value of Q in the delay-state concatenation method. Figure 6 (a) shows the normalized mean-squared errors (NMSEs) for the sixth-order and eighth-order Hénon-map tasks for various values of Q with $P = 1$. As the value of Q increases, the NMSE first decreases, but abruptly

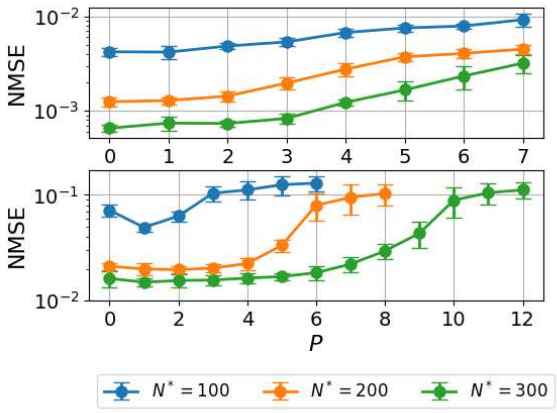

Figure 7: **P-dependence of the learning performance.** NMSEs for the NARMA5 (the top panel) and NARMA10 (the bottom panel) tasks with delay-state concatenation as functions of P for various values of N^*. Error bars show the standard deviation of the results of 10 trials.

increases when Q is larger than 4 in the sixth-order Hénon-map and larger than 6 in the eighth-order Hénon-map. Considering the fact that to predict the output, the mth-order Hénon-map has two informative inputs at the mth and $(m-1)$th previous steps, these increases in performance are reasonable because a RC system with the appropriate values of Q can possess the information needed from past inputs. We also note that similar results were obtained recently in [47]. In contrast, the NMSE monotonically increases as the value of Q increases for NARMA5 and NARMA10, as shown in Fig. 6 (b). These results imply that simply adjusting the value of Q is not effective for tasks such as NARMA5 and NARMA10 with complicated temporal structures.

We next investigated the effects of the value of P for the NARMA tasks in which adjusting the value of Q was not effective. Figure 7 shows the NMSE against the values of P for the NARMA5 task (the top panel) and NARMA10 task (the bottom panel) for three settings of N^*. Note that N^* is fixed as P is varied to ensure that the size of reservoir N^{res} reduces to approximately $1/(P+1)$ times as P varies. We found that the NMSEs were almost constant up to specific values of P. For example, for the case of the NARMA10 task with $N^* = 300$, the NMSE was almost constant up to $P = 5$, indicating that the number of neurons in the reservoir can be reduced from 300 to 50 without impairing performance.

Finally, we compared the proposed methods, delay-state concatenation with and without one transient state, and drift-state concatenation on the NARMA10 task. Figure 8 shows the NMSE as functions of P values with $N^* = 200$ and 300 for the proposed methods. We found that the NMSEs for delay-state concatenation with one transient state and drift-state concatenation were lower than those for delay-state concatenation when P was larger than 7. For the NARMA10 task, inputs more than 10 steps in the past are not very informative for predicting one step forward. The lack of information in the past steps may explain the increase in NMSE for delay-state concatenation when P is larger than 7. The observed lower NMSEs for delay-state concatenation with one transient state and drift-state concatenation are also reasonable because (I) the former method uses the past states of the reservoir, which contain more recent input information for the same value of P (see Fig. 2 D), and (II) the latter method uses the states of the reservoir, which contain current input information.

Discussion

In this study, we proposed three methods to reduce the size of an RC reservoir without impairing performance.

To elucidate the mechanism of the proposed methods, we analyzed the IPC. We found that the value

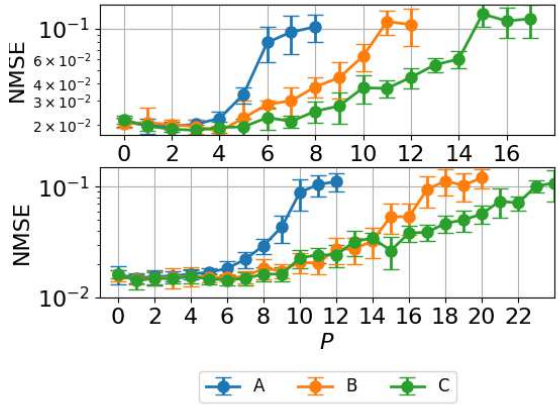

Figure 8: **Comparison of the proposed methods.** Results of regression performance on the NARMA10 task for various values of P with $N^* = 200$ (the top panel) and $N^* = 300$ (the bottom panel) for (A) delay-state concatenation, (B) delay-state concatenation with one transient state, and (C) drift-state concatenation. Error bars show the standard deviation of the results of 10 trials.

of the total IPCs almost reaches $N^{\mathrm{res}}(P+1)$ using the proposed methods, whereas the importance of their components (the first-, third-, and fifth-order IPCs) changes drastically. We also found that the delay structures of the IPCs depend on the value of Q and P. To investigate the applicability of the proposed methods on complex data, we presented the experimental results on generalized Hénon-map and NARMA tasks. We found that when the target task has a relatively simple temporal structure, as demonstrated with the Hénon-map tasks, selecting an appropriate value of Q enhances the performance substantially. In contrast, when the target task contains complex temporal structure, as demonstrated in the NARMA tasks, adjusting the value of Q does not enhance the performance. However, in those cases, we found that increasing the value of P can reduce the size of the reservoir without impairing performance. We have demonstrated that the number of neurons in the reservoir can be reduced by up to one tenth in the NMARMA10 task.

Here, we briefly note the relationship between our work and the most relevant previous work [47]. In [47], the authors proposed a method that is similar to delay-state concatenation. Their proposed model corresponds to the case when the number of additional connected past states is one (i.e., $P = 1$). They observed that performance enhancement depends on the value of Q, as we showed in this paper. However, to the best of our knowledge, the dependence of the performance on the value of P has not been reported. In addition, the other two proposed methods, drift-state concatenation and delay-state concatenation with transient states, are introduced for the first time in this paper. Moreover, the authors of [47] explained the mechanism of their proposed method in terms of the delayed embedding theorem [48]. In contrast, we have provided a more intuitive explanation based on the IPC [38].

Because the proposed methods do not assume a specific topology for the reservoir, they can readily be implemented in FPGAs and physical reservoir systems, such as photonic reservoirs [37]. Therefore, the proposed methods could be an important set of techniques that facilitates the introduction of RC in edge computing.

Methods

Training output weights.

The training procedures are the same as those for standard RC models [7]. The output weights are trained by minimizing

$$\sum_{t=1}^{T} ||\boldsymbol{y}(t) - \boldsymbol{y}^{\mathrm{tc}}(t)||_2^2. \tag{12}$$

Adding a regularization term $\|\hat{W}^{\mathrm{out}}\|_2^2$ did not improve the performance in our case.

Delay-state concatenation with transient states.

We inserted transient states in the RC system with delay-state concatenation as follows:

$$\boldsymbol{x}(t) = \tanh\left(W^{\mathrm{res}}\boldsymbol{x}(t-1) + W^{\mathrm{in}}\boldsymbol{u}\left(\left\lfloor \frac{t}{N^{\mathrm{tran}}+1} \right\rfloor\right)\right), \tag{13}$$

$$\hat{\boldsymbol{x}}(t) = \begin{pmatrix} \boldsymbol{x}(t) \\ \boldsymbol{x}(t-Q) \\ \vdots \\ \boldsymbol{x}(t-QP) \end{pmatrix}, \tag{14}$$

$$\boldsymbol{y}(t) = W^{\mathrm{out}}\hat{\boldsymbol{x}}\left((N^{\mathrm{tran}}+1)t + N^{\mathrm{tran}}\right), \tag{15}$$

where N^{tran} is the number of inserted transient states.

IPC.

Following the same procedure given in [38], the IPCs are calculated as follows: The total IPC is defined as

$$C^{\mathrm{total}} = \sum_{\{d_i\}} C\left(\{d_i\}\right), \tag{16}$$

where C^{total} is the total IPC and $C(\{d_i\})$ is the IPC for a basis represented with a list $\{d_i\} = \{d_0, d_1, \dots\}$. The list represents an orthogonal basis in Hilbert space. We employed the following Legendre polynomials as the orthogonal basis:

$$y_{\{d_i\}}(t) = \prod_{i=0}^{\tau_{\max}} P_{d_i}\left(u(t-i)\right), \tag{17}$$

where $P_{d_i}(\cdot)$ is the d_ith-order Legendre polynomial and $u(t)$ is drawn from uniform distribution on $[-1, 1]$. A constant $\tau_{\max}$ is the maximum delay, which must be large enough to converge the calculation. In our simulations, we set $\tau_{\max}$ to 50 for $\rho^{\mathrm{in}} = 0.1$ and to 25, otherwise. Then, the IPC $C(\{d_i\})$ can be calculated as

$$C(\{d_i\}) = 1 - \frac{\left\langle |y(t) - y_{\{d_i\}}(t)|^2 \right\rangle}{\left\langle |y(t) - \langle y(t)\rangle|^2 \right\rangle}, \tag{18}$$

where $\langle x(t)\rangle := \frac{1}{T}\sum_{t=1}^{T} x(t)$. We set the simulation steps T to 10^6 in all experiments. To avoid overestimation, the value of $C(\{d_i\})$ was set to zero when the value is less than threshold of $7N^{\mathrm{res}}(P+1) \times 10^{-5}$.

We define the kth-order IPCs as

$$C^{k\mathrm{th}} = \sum_{\{d_i\}\in\Gamma^k} C(\{d_i\}), \tag{19}$$

$$\Gamma^k = \left\{ \{d_i\} \middle| \sum_{i=0}^{\tau_{\max}} d_i = k \right\}. \tag{20}$$

The kth-order IPC was decomposed into components corresponding to a subset of a basis whose maximum delay is τ as follows:

$$C^{k\mathrm{th}} = \sum_{\tau=0}^{\tau_{\max}} C_\tau^{k\mathrm{th}}, \tag{21}$$

$$C_\tau^{k\mathrm{th}} = \sum_{\{d_i\}\in\Gamma_\tau^k} C(X, \{d_i\}), \tag{22}$$

$$\Gamma_\tau^k = \left\{ \{d_i\} \middle| \sum_{i=0}^{\tau_{\max}} d_i = k, \max\{i|d_i \geq 1\} = \tau \right\}. \tag{23}$$

Dataset.

The mth-order generalized Hénon-map [49] is given by

$$y^{\mathrm{tc}}(t) = 1.76 - y^{\mathrm{tc}}(t-m+1)^2 - 0.1y^{\mathrm{tc}}(t-m) + \sigma(t) \ (m \geq 2), \tag{24}$$

where σ is Gaussian noise with zero mean and standard deviation of 0.05. The inputs and outputs of the RC are the time series of an n-dimensional generalized Hénon-map. The task is to predict one step forward $y(t+1)$ with past inputs $y(t), y(t-1), \ldots$. The NARMA time series is obtained with a nonlinear auto-regressive moving average as follows:

$$\begin{aligned} y^{\mathrm{tc}}(t) = {} & 0.3y^{\mathrm{tc}}(t-1) + 0.05y^{\mathrm{tc}}(t-1)\sum_{i=1}^{m} y^{\mathrm{tc}}(t-i) \\ & + 1.5s(t-9)s(t) + 0.1, \end{aligned} \tag{25}$$

where $s(t)$ is drawn from uniform distribution of $[0, 0.5]$. The NARMA5 and NARMA10 time series correspond to the case when $m = 5$ and $m = 10$, respectively. The inputs of the RC are $s(t)$. The task is to predict $y(t)$ from the inputs $s(t)$.

For both tasks, we used 2,000 steps as a training dataset and used 3,000 steps as a test dataset. We removed the first 200 steps (free run) both during the training and test phases to avoid the effects of the initial conditions in the reservoirs [7]. We evaluated the performance based on the normalized mean-squared error (NMSE) during the test phase following:

$$\mathrm{NMSE} = \frac{\left\langle |y(t) - y^{\mathrm{tc}}(t)|^2 \right\rangle}{\left\langle |y(t) - \langle y(t) \rangle|^2 \right\rangle}, \tag{26}$$

where $\langle x(t) \rangle = \frac{1}{T}\sum_{t=1}^{T} x(t)$. We averaged the NMSEs over 10 trials. For each iteration, the dataset and connection matrix of the reservoir were generated using their corresponding probabilistic distributions.

Data availability.

All data are generated with computer code.

Computer code.

Computer code is available from the corresponding author on request.

References

[1] K. Greff, R. K. Srivastava, J. Koutnk, B. R. Steunebrink, and J. Schmidhuber. LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 28, No. 10, pp. 2222–2232, Oct 2017.

[2] P. J. Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, Vol. 78, No. 10, pp. 1550–1560, Oct 1990.

[3] Timothy P Lillicrap and Adam Santoro. Backpropagation through time and the brain. *Current Opinion in Neurobiology*, Vol. 55, pp. 82–89, 2019.

[4] Herbert Jaeger. The echo state approach to analysing and training recurrent neural networks. *Technical Report GMD Report 148, German National Research Center for Information Technology*, 2001.

[5] Wolfgang Maass, Thomas Natschlger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, Vol. 14, No. 11, pp. 2531–2560, 2002.

[6] Mantas Lukoeviius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, Vol. 3, No. 3, pp. 127 – 149, 2009.

[7] Mantas Lukoševičius. A practical guide to applying echo state networks. *Neural Networks: Tricks of the Trade*, pp. 659–686, 2012.

[8] Simone Scardapane and Dianhui Wang. Randomness in neural networks: an overview. *WIREs Data Mining Knowl Discov*, Vol. 7, No. 2, p. e1200, 2017.

[9] N. Soures, C. Merkel, D. Kudithipudi, C. Thiem, and N. McDonald. Reservoir computing in embedded systems: Three variants of the reservoir algorithm. *IEEE Consumer Electronics Magazine*, Vol. 6, No. 3, pp. 67–73, July 2017.

[10] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, Vol. 304, No. 5667, pp. 78–80, 2004.

[11] Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.*, Vol. 120, p. 024102, Jan 2018.

[12] Jaideep Pathak, Alexander Wikner, Rebeckah Fussell, Sarthak Chandra, Brian R. Hunt, Michelle Girvan, and Edward Ott. Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos*, Vol. 28, No. 4, p. 041101, 2018.

[13] Patrick L. McDermott and Christopher K. Wikle. An ensemble quadratic echo state network for non-linear spatio-temporal forecasting. *Stat*, Vol. 6, No. 1, pp. 315–330, 2017.

[14] Zhongda Tian, Gang Wang, and Yi Ren. Short-term wind speed forecasting based on autoregressive moving average with echo state network compensation. *Wind Engineering*, Vol. 44, pp. 152–167, 2020.

[15] Xiaowei Lin, Zehong Yang, and Yixu Song. Short-term stock price prediction based on echo state networks. *Expert Systems with Applications*, Vol. 36, No. 3, Part 2, pp. 7313 – 7317, 2009.

[16] Chi-Yi Tsai, Xavier Dutoit, Kai-Tai Song, Hendrik Van Brussel, and Marnix Nuttin. Robust face tracking control of a mobile robot using self-tuning Kalman filter and echo state network. *Asian Journal of Control*, Vol. 12, No. 4, pp. 488–509, 2010.

[17] E. A. Antonelo and B. Schrauwen. On learning navigation behaviors for small mobile robots with reservoir computing architectures. *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 26, No. 4, pp. 763–780, April 2015.

[18] A. Jalalvand, G. V. Wallendael, and R. V. D. Walle. Real-time reservoir computing network-based systems for detection tasks on visual contents. In *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*, pp. 146–151, June 2015.

[19] Pieter Buteneers, David Verstraeten, Bregt Van Nieuwenhuyse, Dirk Stroobandt, Robrecht Raedt, Kristl Vonck, Paul Boon, and Benjamin Schrauwen. Real-time detection of epileptic seizures in animal models using reservoir computing. *Epilepsy Research*, Vol. 103, No. 2, pp. 124 – 134, 2013.

[20] Priyadarshini Panda and Narayan Srinivasa. Learning to recognize actions from limited training examples using a recurrent spiking neural model. *Frontiers in Neuroscience*, Vol. 12, p. 126, 2018.

[21] Lennert Appeltant, Miguel Cornelles Soriano, Guy Van der Sande, Jan Danckaert, Serge Massar, Joni Dambre, Benjamin Schrauwen, Claudio R Mirasso, and Ingo Fischer. Information processing using a single dynamical node as complex system. *Nature communications*, Vol. 2, p. 468, 2011.

[22] Mustafa C Ozturk, Dongming Xu, and José C Príncipe. Analysis and design of echo state networks. *Neural computation*, Vol. 19, No. 1, pp. 111–138, 2007.

[23] A. Rodan and P. Tino. Minimum complexity echo state network. *IEEE Transactions on Neural Networks*, Vol. 22, No. 1, pp. 131–144, Jan 2011.

[24] Tobias Strauss, Welf Wustlich, and Roger Labahn. Design strategies for weight matrices of echo state networks. *Neural computation*, Vol. 24, No. 12, pp. 3246–3276, 2012.

[25] Y. Katori, H. Tamukoh, and T. Morie. Reservoir computing based on dynamics of pseudo-billiard system in hypercube. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, July 2019.

[26] M. L. Alomar, M. C. Soriano, M. Escalona-Morn, V. Canals, I. Fischer, C. R. Mirasso, and J. L. Rossell. Digital implementation of a single dynamical node reservoir computer. *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 62, No. 10, pp. 977–981, Oct 2015.

[27] Lisa Loomis, Nathan McDonald, and Cory Merkel. An FPGA implementation of a time delay reservoir using stochastic logic. *J. Emerg. Technol. Comput. Syst.*, Vol. 14, No. 4, p. 46, December 2018.

[28] M. L. Alomar, Erik S. Skibinsky-Gitlin, Christiam F. Frasser, Vincent Canals, Eugeni Isern, Miquel Roca, and Josep L. Rosselló. Efficient parallel implementation of reservoir computing systems. *Neural Computing and Applications*, Vol. 32, pp. 2299–2313, Dec 2018.

[29] Bogdan Penkovsky, Laurent Larger, and Daniel Brunner. Efficient design of hardware-enabled reservoir computing in FPGAs. *Journal of Applied Physics*, Vol. 124, No. 16, p. 162101, 2018.

[30] D. Brunner, B. Penkovsky, B. A. Marquez, M. Jacquot, I. Fischer, and L. Larger. Tutorial: Photonic neural networks in delay systems. *Journal of Applied Physics*, Vol. 124, No. 15, p. 152004, 2018.

[31] Thomas Ferreira de Lima, Bhavin J Shastri, Alexander N Tait, Mitchell A Nahmias, and Paul R Prucnal. Progress in neuromorphic photonics. *Nanophotonics*, Vol. 6, No. 3, pp. 577–599, 2017.

[32] H. Peng, M. A. Nahmias, T. F. de Lima, A. N. Tait, and B. J. Shastri. Neuromorphic photonic integrated circuits. *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 24, No. 6, pp. 1–15, Nov 2018.

[33] Jacob Torrejon, Mathieu Riou, Flavio Abreu Araujo, Sumito Tsunegi, Guru Khalsa, Damien Querlioz, Paolo Bortolotti, Vincent Cros, Kay Yakushiji, Akio Fukushima, et al. Neuromorphic computing with nanoscale spintronic oscillators. *Nature*, Vol. 547, pp. 428–431, 2017.

[34] Guillaume Dion, Salim Mejaouri, and Julien Sylvestre. Reservoir computing with a single delay-coupled non-linear mechanical oscillator. *Journal of Applied Physics*, Vol. 124, No. 15, p. 152132, 2018.

[35] F. C. Bauer, D. R. Muir, and G. Indiveri. Real-time ultra-low power ECG anomaly detection using an event-driven neuromorphic processor. *IEEE Transactions on Biomedical Circuits and Systems*, Vol. 13, No. 6, pp. 1575–1582, 2019.

[36] M. Yamaguchi, Y. Katori, D. Kamimura, H. Tamukoh, and T. Morie. A chaotic Boltzmann machine working as a reservoir and its analog VLSI implementation. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, July 2019.

[37] Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Hroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. Recent advances in physical reservoir computing: A review. *Neural Networks*, Vol. 115, pp. 100 – 123, 2019.

[38] Joni Dambre, David Verstraeten, Benjamin Schrauwen, and Serge Massar. Information processing capacity of dynamical systems. *Scientific reports*, Vol. 2, p. 514, 2012.

[39] Herbert Jaeger. Short term memory in echo state networks. *Technical Report GMD Report 152, German National Research Center for Information Technology*, 2002.

[40] Olivia L. White, Daniel D. Lee, and Haim Sompolinsky. Short-term memory in orthogonal neural networks. *Phys. Rev. Lett.*, Vol. 92, p. 148102, Apr 2004.

[41] Joschka Boedecker, Oliver Obst, Joseph T. Lizier, N. Michael Mayer, and Minoru Asada. Information processing in echo state networks at the edge of chaos. *Theory in Biosciences*, Vol. 131, No. 3, pp. 205–213, Sep 2012.

[42] Igor Farkaš, Radomr Bosk, and Peter Gerge. Computational analysis of memory capacity in echo state networks. *Neural Networks*, Vol. 83, pp. 109 – 120, 2016.

[43] Surya Ganguli, Dongsung Huh, and Haim Sompolinsky. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, Vol. 105, No. 48, pp. 18970–18975, 2008.

[44] Izzet B. Yildiz, Herbert Jaeger, and Stefan J. Kiebel. Re-visiting the echo state property. *Neural Networks*, Vol. 35, pp. 1–9, 2012.

[45] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems 25*, pp. 2951–2959. 2012.

[46] Peter I Frazier. A tutorial on bayesian optimization. *arXiv:1807.02811*, 2018.

[47] Bicky A. Marquez, Jose Suarez-Vargas, and Bhavin J. Shastri. Takens-inspired neuromorphic processor: A downsizing tool for random recurrent neural networks via feature extraction. *Phys. Rev. Research*, Vol. 1, p. 033030, Oct 2019.

[48] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*. Springer, Berlin, 1981.

[49] Hendrik Richter. The generalized Hénon maps: Examples for higher-dimensional chaos. *International Journal of Bifurcation and Chaos*, Vol. 12, No. 06, pp. 1371–1384, 2002.

Acknowledgment

The authors would like to thank Makoto Ikeda, Hiromitsu Awano, and Gouhei Tanaka for the fruitful discussion. This work was partially supported by the "Brain-Morphic AI to Resolve Social Issues" project at UTokyo, the NEC Corporation, and AMED (JP20dm0307009).

Author contributions

All the authors designed the research. Y.S performed all simulations, and all the authors confirmed the theory. Further, they all wrote the paper.

Competing interests

Y.S., K.M., and K.A. have applied for a patent related to the proposed methods (Japanese Patent Application No. 2019-206438). The remaining author has no conflict of interest to declare.

Figures

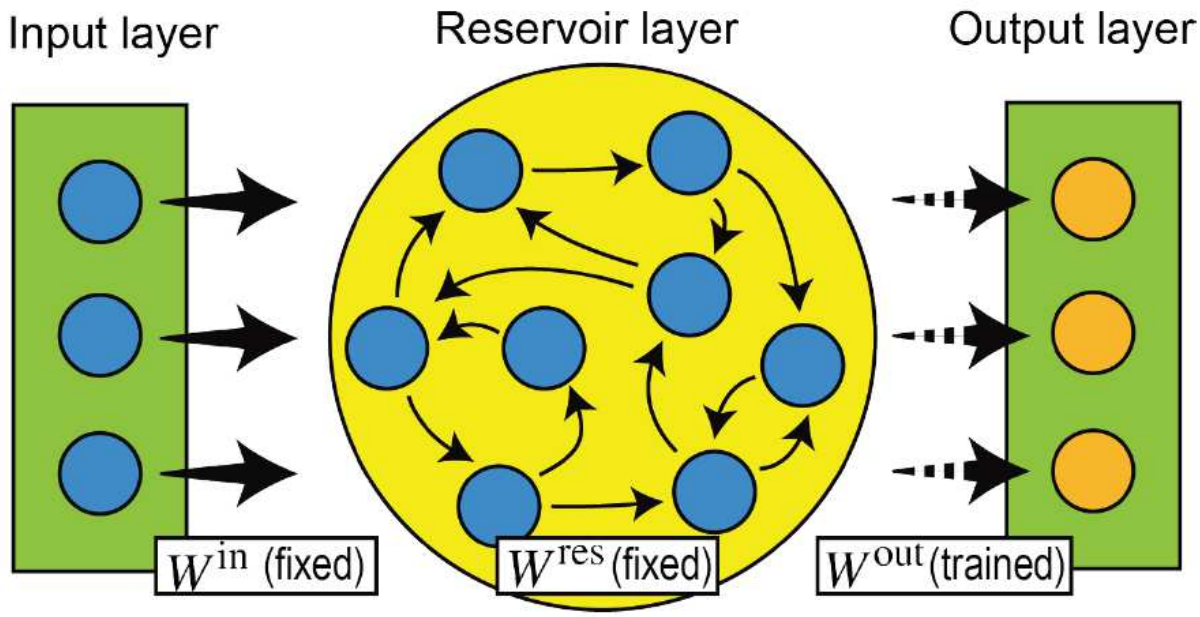

Figure 1

Typical RC architecture. The reservoir layer consists of randomly connected neurons. The connections between the input and reservoir layer Win and connections within the reservoir layer Wres are fixed (solid arrows), whereas the output weights Wout are trained (dashed arrows).

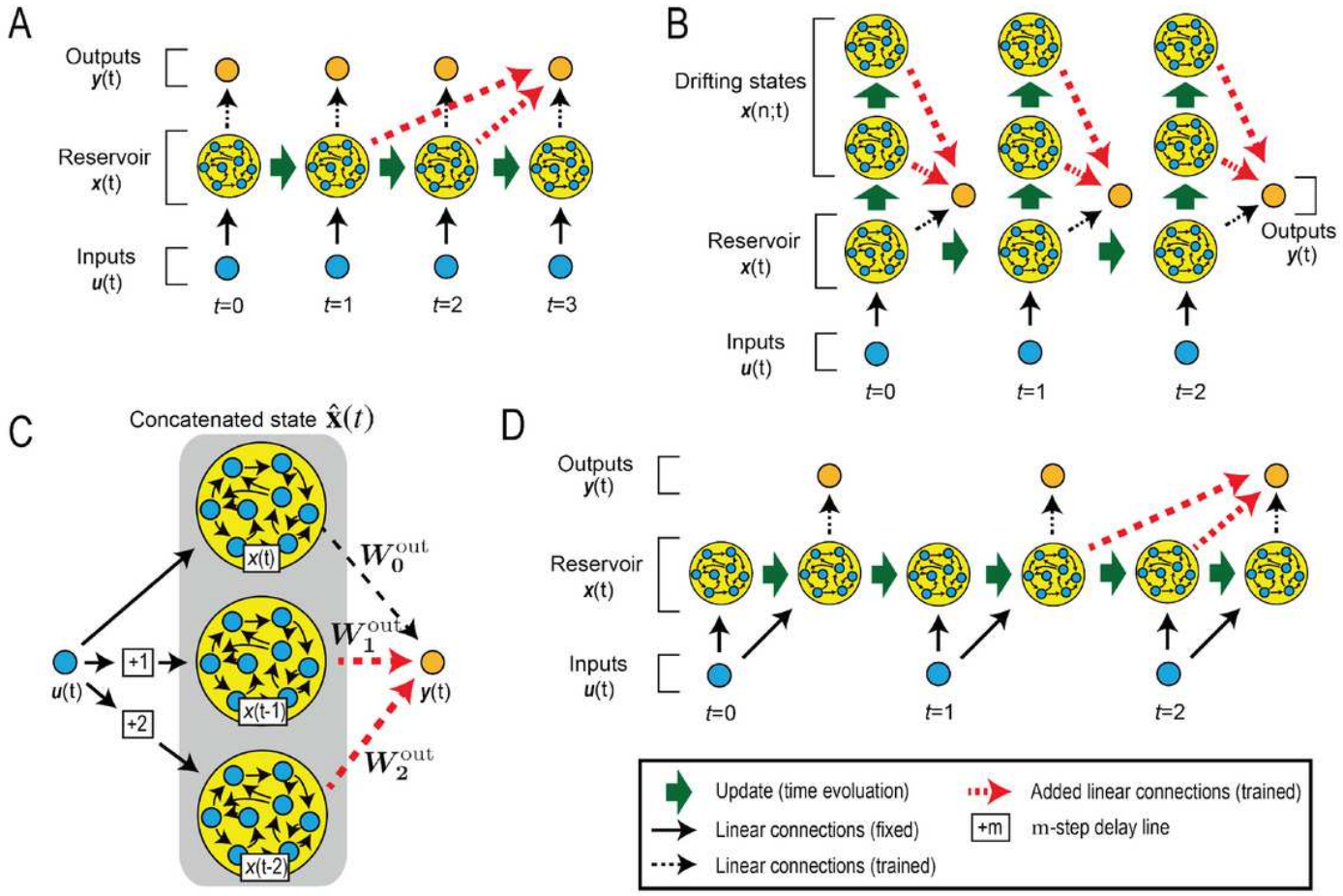

Figure 2

Schematics of the proposed methods. A. Delay-state concatenation when the number of additional connections P is two and the unit of delay Q is one. B. Drift-state concatenation when P is two and Q is one. C. Another view of delay-state concatenation. The reservoir consists of three identical dynamical systems and delay lines. The added dynamical systems have +1 delay lines and +2 delay lines, respectively. D. Delay-state concatenation with one transient state when P is two and Q is one.

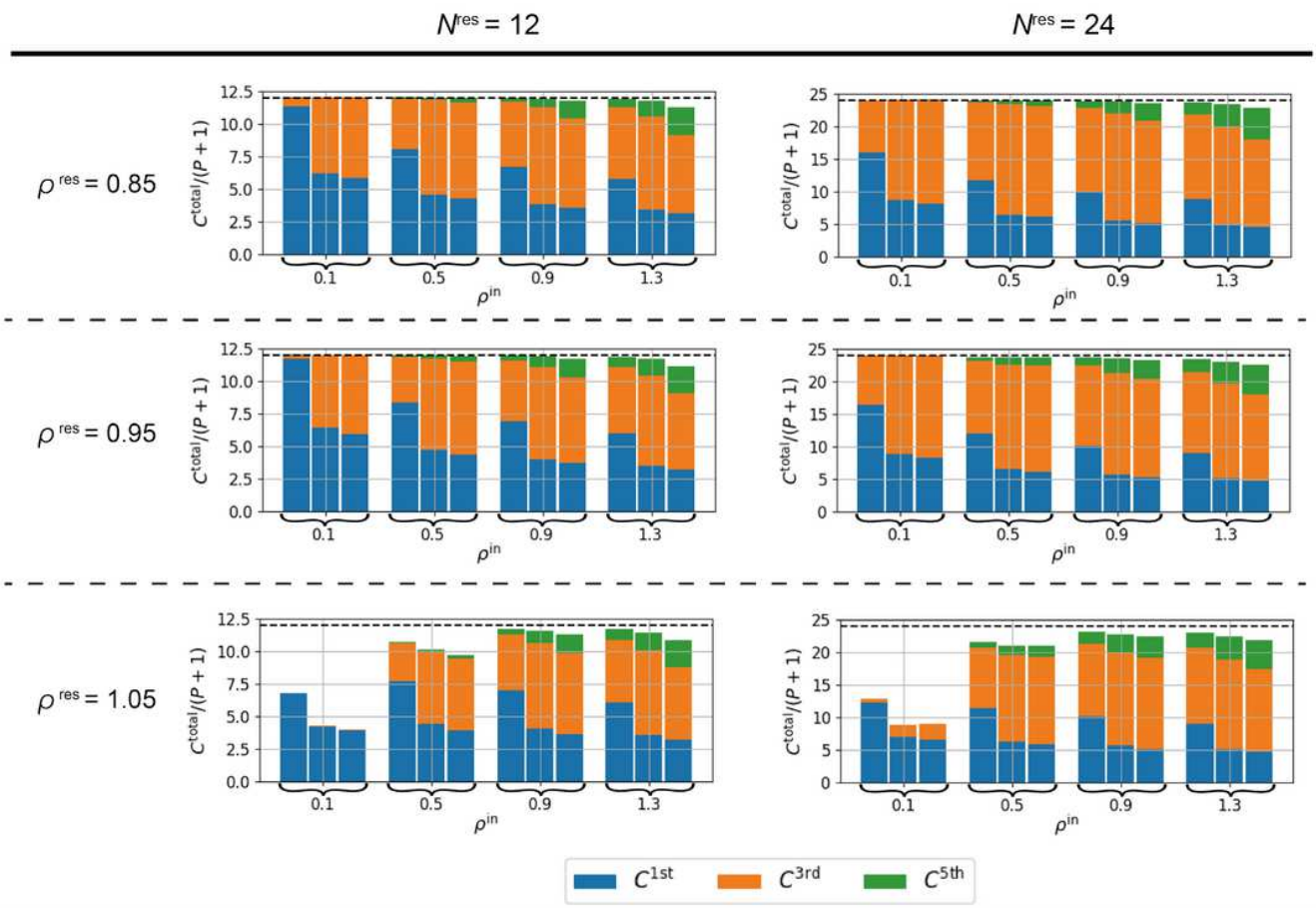

Figure 3

Analysis of information processing capacities (IPCs). IPCs of standard RC and RC with the proposed methods for various values of input weight strength in and spectral radius res. The left and right panels show the results for Nres = 12 and Nres = 24 , respectively. The dashed horizontal line within each subgraph represents the value of Nres. For each value of in, the left column presents the standard RC with P = 0, the center column presents delay-state concatenation with P = Q = 1, and the right column presents drift-state concatenation with P = Q = 1.

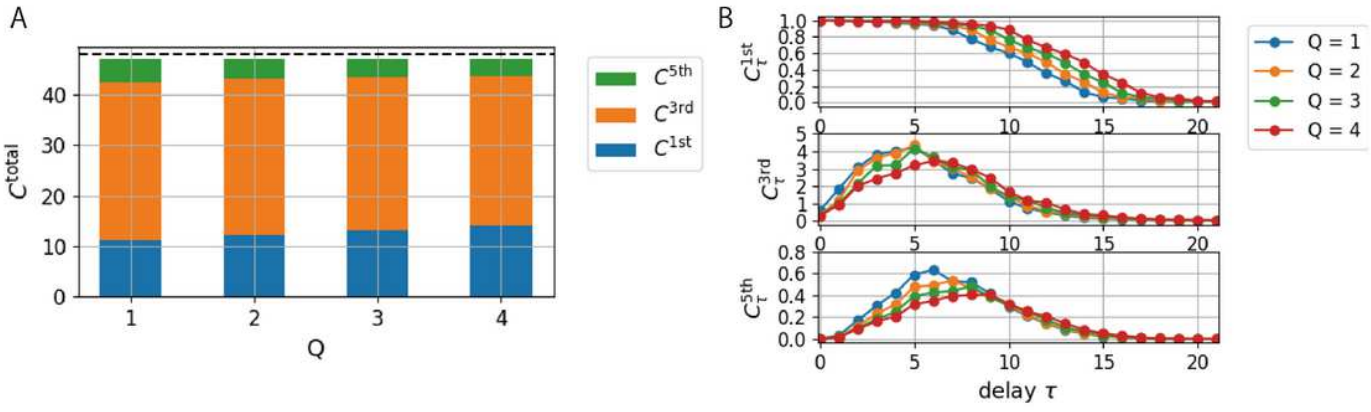

Figure 4

Q-dependence of IPCs. (A) IPCs of the RC with delay-state concatenation for various values of Q. (B) Delay structures of the IPCs for various values of Q. From top to bottom, the first-, third-, and fifth-order IPCs are shown. The parameters are set as follows: Nres = 24, pin = 0:9, pres = 0:95, and P = 1.

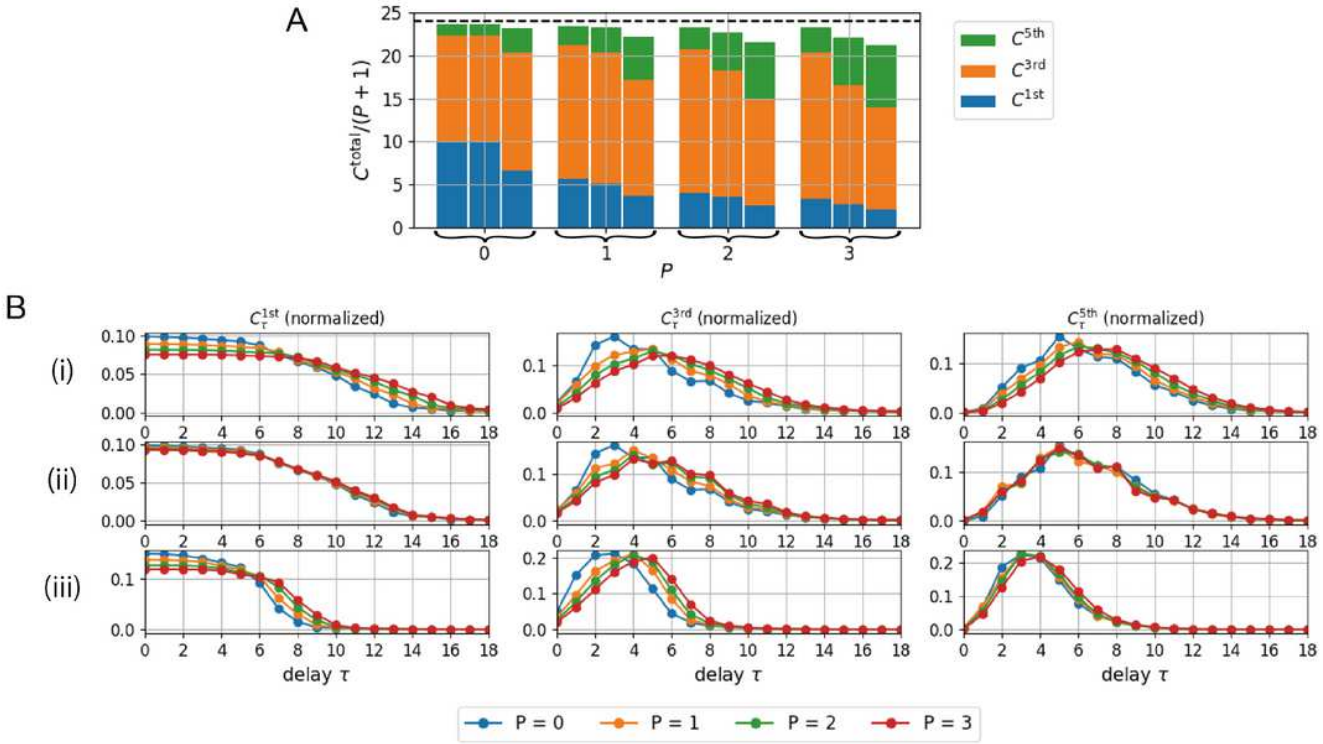

Figure 5

P-dependence of IPCs. (A) IPCs with the proposed methods for various values of P. Left columns: IPCs for delay-state concatenation, center columns: IPCs for drift-state concatenation, and right columns: IPCs for delay-state concatenation with one transient state. (B) Comparison of the delay structures of the IPCs for the three proposed methods. Top panels (i): delay structures for delay-state concatenation, middle panels (ii): delay structures for drift-state concatenation, and bottom panels (iii): delay structures for delay-state concatenation with one transient state. To highlight the difference in the distribution of the delay structures, the IPCs for each order are normalized. From left to right, the first-, third-, and fifth-order IPCs are shown. The parameters are set as follows: Nres = 24, pin = 0:9, pres = 0:95, and Q = 1.

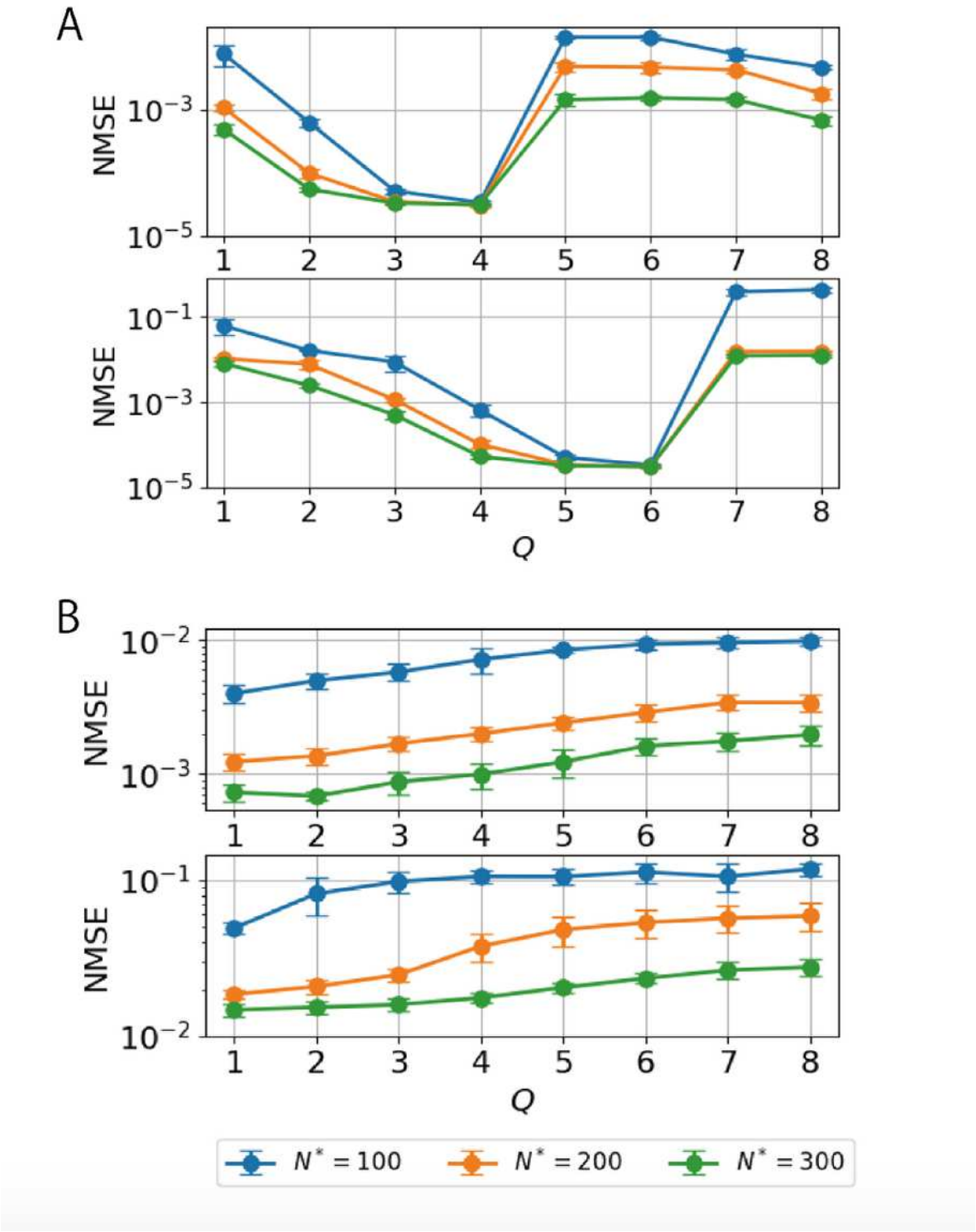

Figure 6

Q-dependence of the learning performance. (A) NMSE for a sixth-order Hénon-map task (the upper panel) and for an eighth-order Hénon-map task (the lower panel) for various values of Q with P = 1. (B) NMSE for NARMA5 (the upper panel) task and for NARMA10 (the lower panel) task for various values of Q with P = 1. Error bars show the standard deviation of the results of 10 trials.

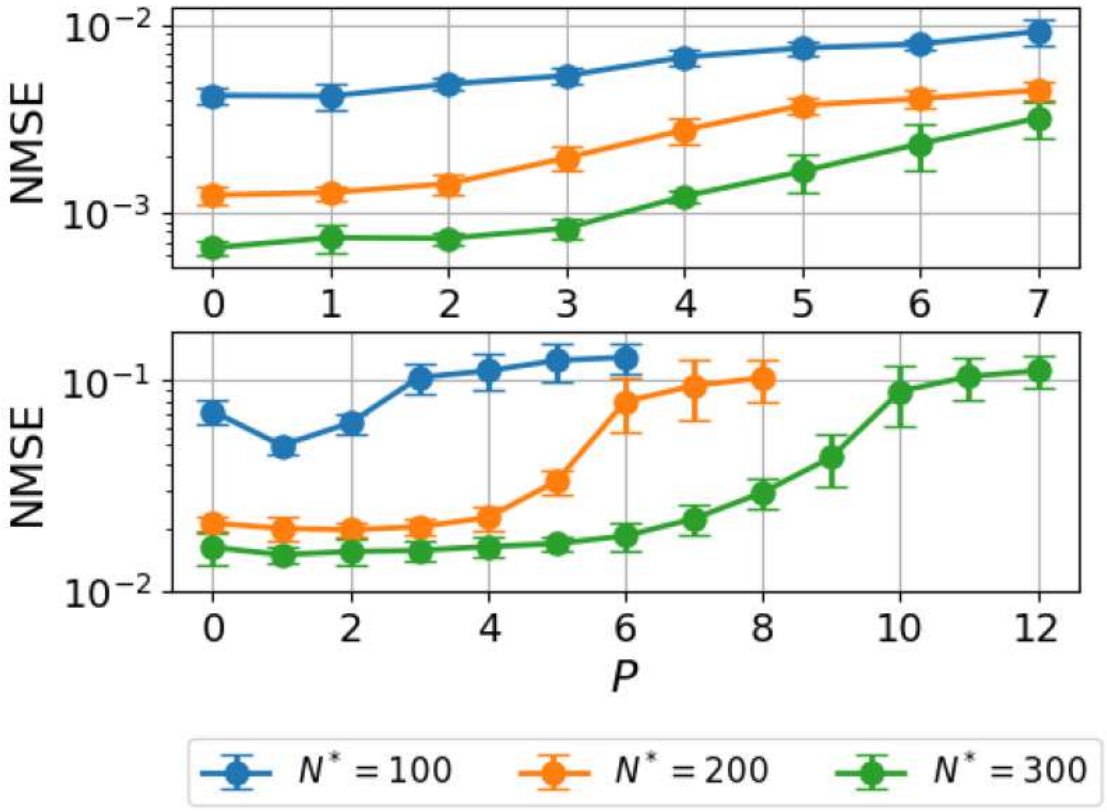

Figure 7

P-dependence of the learning performance. NMSEs for the NARMA5 (the top panel) and NARMA10 (the bottom panel) tasks with delay-state concatenation as functions of P for various values of N*. Error bars show the standard deviation of the results of 10 trials.

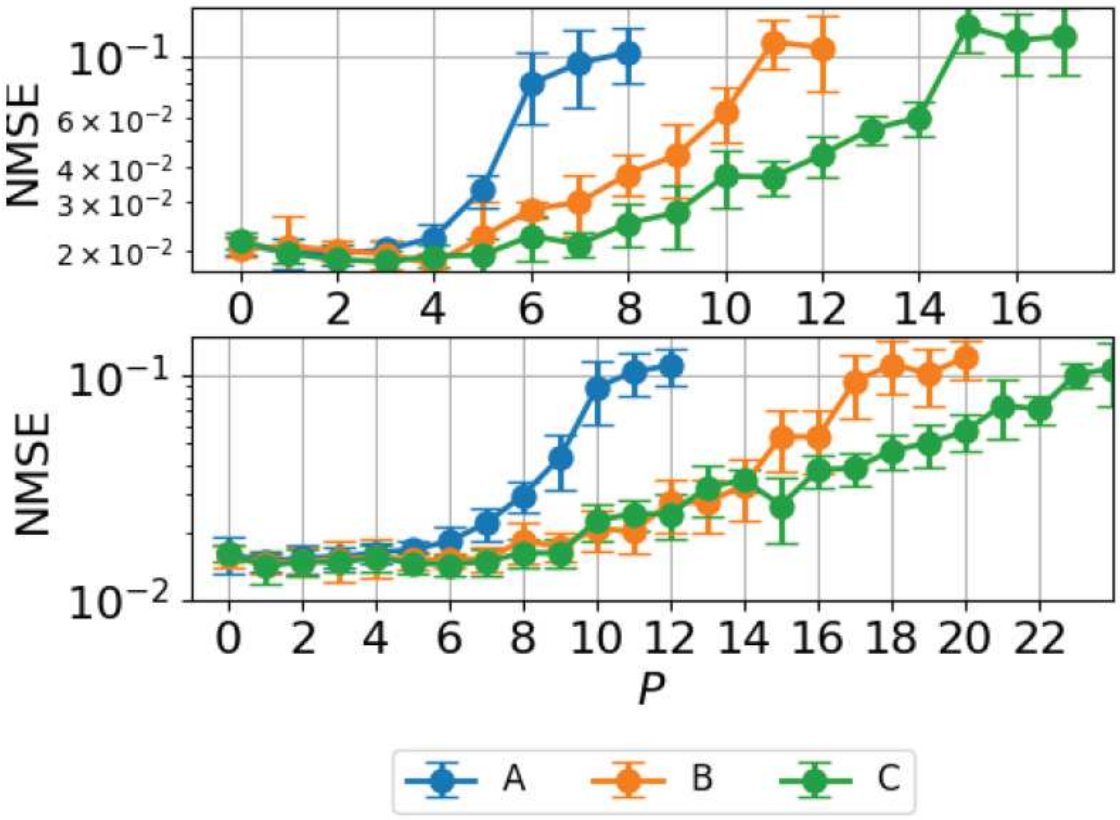

Figure 8

Comparison of the proposed methods. Results of regression performance on the NARMA10 task for various values of P with N* = 200 (the top panel) and N* = 300 (the bottom panel) for (A) delay-state concatenation, (B) delay-state concatenation with one transient state, and (C) drift-state concatenation. Error bars show the standard deviation of the results of 10 trials.