

Reservoir computing with diverse timescales for prediction of multiscale dynamics

Gouhei Tanaka ^{1,2,3}, Tadayoshi Matsumori ⁴, Hiroaki Yoshida ⁴, and Kazuyuki Aihara ¹¹*International Research Center for Neurointelligence, The University of Tokyo, Tokyo 113-0033, Japan*²*Department of Electrical Engineering and Information Systems, Graduate School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan*³*Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan*⁴*Toyota Central R&D Labs., Inc., Tokyo 112-0004, Japan*

(Received 23 August 2021; accepted 28 June 2022; published 28 July 2022)

Machine learning approaches have recently been leveraged as a substitute or an aid for physical/mathematical modeling approaches to dynamical systems. To develop an efficient machine learning method dedicated to modeling and prediction of multiscale dynamics, we propose a reservoir computing (RC) model with diverse timescales by using a recurrent network of heterogeneous leaky integrator (LI) neurons. We evaluate computational performance of the proposed model in two time series prediction tasks related to four chaotic fast-slow dynamical systems. In a one-step-ahead prediction task where input data are provided only from the fast subsystem, we show that the proposed model yields better performance than the standard RC model with identical LI neurons. Our analysis reveals that the timescale required for producing each component of target multiscale dynamics is appropriately and flexibly selected from the reservoir dynamics by model training. In a long-term prediction task, we demonstrate that a closed-loop version of the proposed model can achieve longer-term predictions compared to the counterpart with identical LI neurons depending on the hyperparameter setting.

DOI: [10.1103/PhysRevResearch.4.L032014](https://doi.org/10.1103/PhysRevResearch.4.L032014)

Introduction. Hierarchical structures composed of macroscale and microscale components are ubiquitous in physical, biological, medical, and engineering systems [1–4]. Complex interactions between such diverse components often bring about multiscale behavior with different spatial and temporal scales. To understand such complex systems, *multiscale modeling* has been one of the major challenges in science and technology [5]. An effective approach is to combine established physical models at different scales by considering their interactions. However, even a physical model focusing on one scale is often not available when the rules (e.g., physical laws) governing the target system are not fully known. In such a case, another potential approach is to employ a machine learning model fully or partly [6,7]. It is a challenging issue to integrate machine learning and multiscale modeling for dealing with large datasets from different sources and different levels of resolution [8]. To this end, it is significant to develop robust predictive machine learning models specialized for multiscale dynamics.

We focus on a machine learning framework called *reservoir computing* (RC), which has been mainly applied to temporal pattern recognition such as system identification, time series prediction, time series classification, and anomaly

detection [6,9–13]. The echo state network (ESN) [9,10], which is a major RC model, uses a recurrent neural network (RNN) to nonlinearly transform an input time series into a high-dimensional dynamic state and reads out desired characteristics from the dynamic state. The RNN with fixed random connection weights serves as a *reservoir* to generate an echo of the past inputs. Compared to other RNNs where all the connection weights are trained with gradient-based learning rules [14], the ESN can be trained with much lower computational cost by optimizing only the readout parameters with a simple learning method [15].

The timescale of reservoir dynamics in the original ESN [9] is almost determined by that of input time series. However, the timescale of a desired output time series is often largely different from that of the input one depending on a learning task. Therefore, the ESN with leaky integrator neurons (LI-ESN) has been widely used as a standard model to accommodate the model output to temporal characteristics of the target dynamics [16]. The LI-ESN has a leak rate parameter controlling the update speed of the neuronal states in the reservoir. For multi-timescale dynamics, it is an option to use the hierarchical ESN combining multiple reservoirs with different timescales [17,18], where the leak rate is common to all the neurons in each reservoir but can be different from one reservoir to another. In the above models, the leak rate in each reservoir is set at an optimal value through a grid search or a gradient-based optimization.

In contrast to the above approach, here we propose an ESN with diverse timescales (DTS-ESN), where the leak rates of the LI neurons are distributed. Our aim is to generate

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

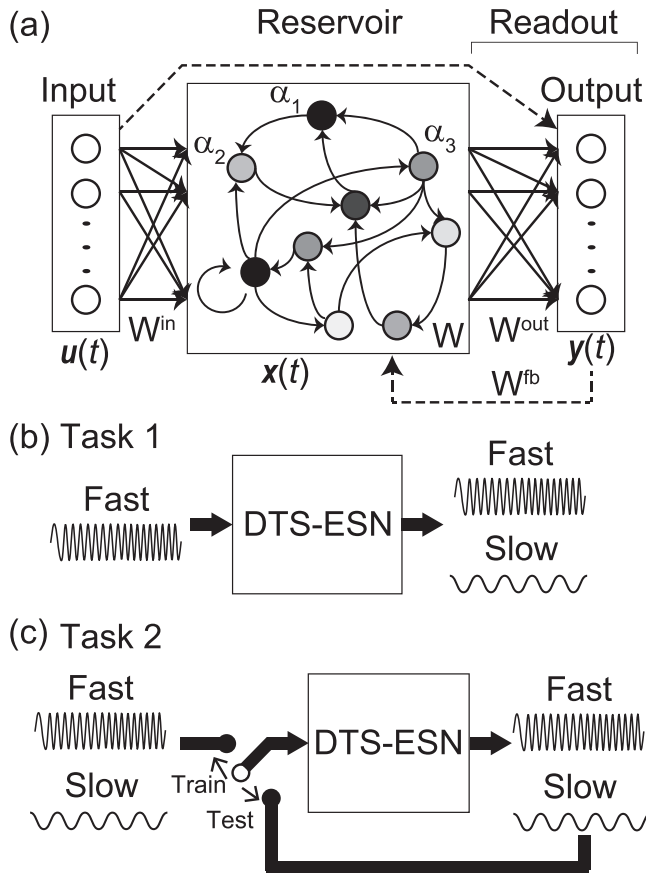


FIG. 1. (a) A schematic of the DTS-ESN. An input time series $\mathbf{u}(t)$ given to the input layer is transformed into a reservoir state $\mathbf{x}(t)$, and then processed in the readout to produce an output time series $\mathbf{y}(t)$. Only W^{out} is trainable. The dashed arrows indicate optional connections. The reservoir neuron i has leak rate $\alpha_i \in [\alpha_{\min}, \alpha_{\max}]$. (b) Task 1 where the whole fast-slow dynamics is predicted only from the fast dynamics using an open-loop model. (c) Task 2 where the whole fast-slow dynamics is predicted with a closed-loop model.

reservoir dynamics with a rich variety of timescales so as to accommodate a wide range of desired time series with different timescales. The main advantage of the DTS-ESN is flexible adjustability to multiscale dynamics. Moreover, the idea behind the proposed model opens up a new possibility to leverage heterogeneous network components for network-type physical reservoirs in implementation of RC hardware [19] as well as for other variants of ESNs and RC models [20–23]. Our idea is also motivated by the role of heterogeneity of biological neurons in information flow and processing [24].

Some previous works considered heterogeneity of system components in RC systems. A positive effect of heterogeneous nonlinear activation functions of reservoir neurons on prediction performance is reported [25], but its explicit relation to timescales of reservoir states is unclear. In a time-delay-based reservoir [26], where the timescales of system dynamics are mainly governed by the delay time and the clock cycle, a mismatch between them can enhance computational performance [27,28]. In a time-delay-based physical reservoir implemented with an optical fiber-ring cavity, it

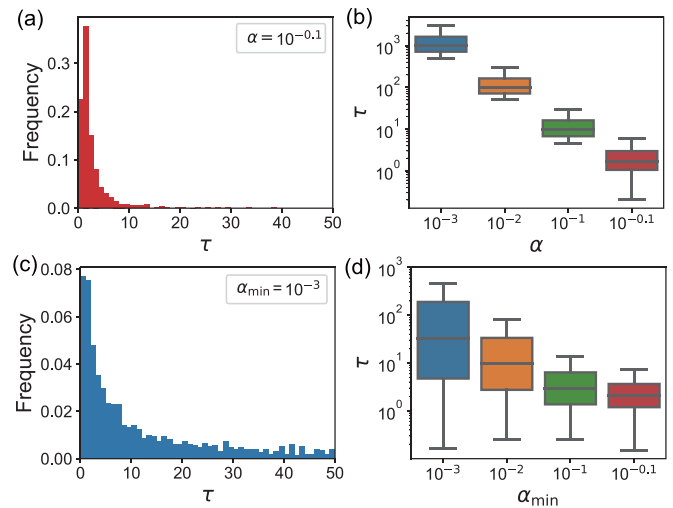


FIG. 2. Timescale distributions of reservoir dynamics in the linearized systems of the LI-ESN and the DTS-ESN when $N_x = 2000$ and $d = 0.1$. (a) A histogram of timescales $[\tau_i$ in Eq. (5)] in the LI-ESN with $\alpha = 10^{-0.1}$. (b) The effect of α on the timescale distribution in the LI-ESN. (c) The same as (a) but in the DTS-ESN with $\alpha_{\max} = 1$ and $\alpha_{\min} = 10^{-3}$. (d) The effect of α_{\min} on the timescale distribution in the DTS-ESN with $\alpha_{\max} = 1$.

was revealed that the nonlinearity of the fiber waveguide is essential for high computational performance rather than the nonlinearity in the input and readout parts [29]. Compared to these time-delay-based reservoirs with a few controllable timescale-related parameters, our model based on a network-type reservoir can realize various timescale distributions by setting different heterogeneity of leak rates.

Methods. The DTS-ESN consists of a reservoir with heterogeneous LI neurons and a linear readout as illustrated in Fig. 1(a). The DTS-ESN receives an input time series $\mathbf{u}(t)$ in the input layer, transforms it into a high-dimensional reservoir state $\mathbf{x}(t)$, and produces an output time series $\mathbf{y}(t)$ as a linear combination of the states of the reservoir neurons. With distributed leak rates of reservoir neurons, the DTS-ESN extends greatly the range of timescales that can be realized by the LI-ESN [16].

In this study, the capability of the DTS-ESN is evaluated in two types of time series prediction tasks related to chaotic fast-slow dynamical systems. One is a one-step-ahead prediction task (Task 1) where input data are given only from the fast subsystem as depicted in Fig. 1(b). An inference of hidden slow dynamics from observational data of the fast dynamics is challenging but beneficial in reducing the effort involved in data measurement. An example in climate science is to predict slowly changing behavior in the deep ocean such as the temperature. The behavior is known to give an important feedback to the fast dynamical behavior of the atmosphere, land, and near-surface oceans [30], but the observation in the deep ocean remains as a major challenge [31]. The other is an autoregressive prediction task (Task 2) where a closed-loop version of the proposed model is used for a long-term prediction in a testing phase as illustrated in Fig. 1(c). RC approaches have shown a strong potential in long-term predictions of chaotic behavior [6,7]. We examine the effect of

TABLE I. Chaotic fast-slow dynamical systems models used for prediction tasks. Each model consists of fast and slow subsystems of the variables specified in the third and fourth columns. The parameter values were set to exhibit chaotic dynamics.

Model	Equations	Fast	Slow	Parameter values
(i) Rulkov [32]	$x_{t+1} = \eta/(1 + x_t^2) + y_t, y_{t+1} = y_t - \mu(x_t - \sigma)$	x	y	$\eta = 4.2, \mu = 0.001, \sigma = -1.0$
(ii) Hindmarsh-Rose [33]	$\dot{x} = y - x^3 + bx^2 - z + I, \dot{y} = 1 - 5x^2 - y,$ $\dot{z} = \epsilon(s(x - x_0) - z)$	x, y	z	$b = 3.05, I = 3, \epsilon = 0.01,$ $s = 4, x_0 = -1.6$
(iii) tc-VdP [34]	$\dot{x}_1 = (y_1 + c_1x_2)/\tau_1, \dot{y}_1 = (\mu_1(1 - x_1^2)y_1 - x_1)/\tau_1$ $\dot{x}_2 = (y_2 + c_2x_1)/\tau_2, \dot{y}_2 = (\mu_2(1 - x_2^2)y_2 - x_2)/\tau_2$	x_1, y_1	x_2, y_2	$\mu_1 = 5, \tau_1 = 0.1, c_1 = 0.001$ $\mu_2 = 5, \tau_2 = 1, c_2 = 1$
(iv) tc-Lorenz [35]	$\dot{X} = a(Y - X), \dot{Y} = r_sX - ZX - Y - \epsilon_sxy,$ $\dot{Z} = XY - bZ, \dot{x} = ca(y - x),$ $\dot{y} = c(r_fx - zx - y) + \epsilon_fYx, \dot{z} = c(xy - bz)$	x, y, z	X, Y, Z	$a = 10, b = 8/3, c = 10$ $r_s = 28, r_f = 45,$ $\epsilon_s = 0.01, \epsilon_f = 10$

distributed timescales of reservoir dynamics on the long-term prediction ability. The DTS-ESN is formulated as follows:

$$\mathbf{x}(t + \Delta t) = (I - A)\mathbf{x}(t) + Af(\mathbf{h}(t)), \quad (1)$$

$$\mathbf{h}(t) = \rho W\mathbf{x}(t) + \gamma W^{\text{in}}\mathbf{u}(t + \Delta t) + \zeta W^{\text{fb}}\mathbf{y}(t), \quad (2)$$

$$\mathbf{y}(t + \Delta t) = \begin{cases} W^{\text{out}}[\mathbf{x}(t + \Delta t); \mathbf{u}(t + \Delta t); 1] & \text{(Task 1)} \\ W^{\text{out}}\mathbf{x}(t + \Delta t) & \text{(Task 2)}, \end{cases} \quad (3)$$

where t is the time, Δt is the time step, $\mathbf{x}(t) \in \mathbb{R}^{N_x}$ is the reservoir state vector, $I \in \mathbb{R}^{N_x \times N_x}$ is the identity matrix, $A = \text{diag}(\alpha_1, \dots, \alpha_{N_x})$ is the diagonal matrix of leak rates α_i for $i = 1, \dots, N_x$, $\mathbf{h}(t)$ is the internal state vector, f is the element-wise activation function given as $f = \tanh$ in this study, $\rho W \in \mathbb{R}^{N_x \times N_x}$ is the reservoir weight matrix with spectral radius ρ , $\gamma W^{\text{in}} \in \mathbb{R}^{N_x \times N_u}$ is the input weight matrix with input scaling factor γ , $\mathbf{u}(t) \in \mathbb{R}^{N_u}$ is the input vector, $\zeta W^{\text{fb}} \in \mathbb{R}^{N_x \times N_y}$ is the feedback weight matrix with feedback scaling factor ζ , $\mathbf{y}(t) \in \mathbb{R}^{N_y}$ is the output vector, and W^{out} is the output weight matrix. In the readout, we concatenate the reservoir state, the input, and the bias for Task 1 and use only the reservoir state for Task 2 as described in Eq. (3). Only W^{out} is trainable and all the other parameters are fixed in advance [9,13]. The DTS-ESN is reduced to the LI-ESN [16] if $\alpha_i = \alpha$ for all i .

The fraction of nonzero elements in W was fixed at $d = 0.1$, which were randomly drawn from a uniform distribution in $[-1, 1]$, and W was rescaled so that its spectral radius equals 1. The entries of W^{in} and W^{fb} were randomly drawn from a uniform distribution in $[-1, 1]$. The leak rates were assumed to follow a reciprocal (or log-uniform) distribution in $[\alpha_{\min}, \alpha_{\max}]$. This means that $\log_{10} \alpha_i$ was randomly drawn from a uniform distribution in $[\log_{10} \alpha_{\min}, \log_{10} \alpha_{\max}]$. The leak rate α_i of neuron i is represented as $\alpha_i = \Delta t/c_i$ where

TABLE II. The time steps and the durations of transient, training, and testing periods.

Model	Δt	T_{trans}	T_{train}	T_{test}
(i) Rulkov	1	4000	8000	4000
(ii) Hindmarsh-Rose	0.05	200	1200	600
(iii) tc-VdP	0.01	50	150	100
(iv) tc-Lorenz	0.01	30	60	30

c_i denotes the time constant in the corresponding continuous-time model (see the Supplemental Material [36]).

In the training phase, an optimal output weight matrix \hat{W}^{out} is obtained by minimizing the following sum of squared output errors plus the regularization term:

$$\sum_k \|\mathbf{y}(t + k\Delta t) - \mathbf{d}(t + k\Delta t)\|_2^2 + \beta \|W^{\text{out}}\|_F^2, \quad (4)$$

where the summation is taken for all discrete time points in the training period, $\|\cdot\|_2$ indicates the L2 norm, $\|\cdot\|_F$ indicates the Frobenius norm, $\mathbf{d}(t)$ denotes the target dynamics, and β represents the regularization factor [13]. In the testing phase, \hat{W}^{out} is used to produce predicted outputs.

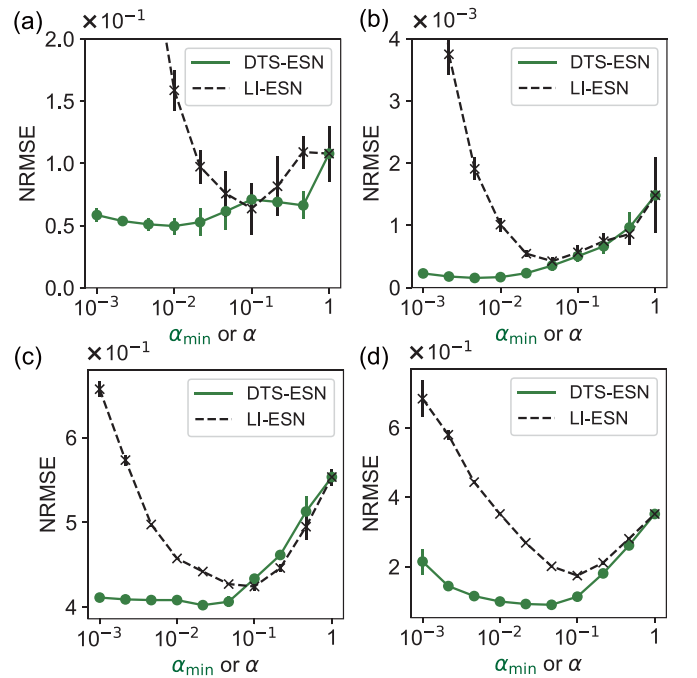


FIG. 3. Comparisons of the testing errors (NRMSEs) between the DTS-ESN with $\alpha_i \in [\alpha_{\min}, 1]$ (green circles) and the LI-ESN with $\alpha_i = \alpha$ (black crosses). The marks indicate the average errors over ten simulation runs with different reservoir realizations. The error bar indicates the variance. The common parameter values are $\zeta = 0$ and $\beta = 10^{-3}$. (a) The Rulkov model. $N_x = 200$ and $\gamma = \rho = 1$. (b) The HR model. $N_x = 200$ and $\gamma = \rho = 1$. (c) The tc-VdP model. $N_x = 200$ and $\gamma = \rho = 1$. (d) The tc-Lorenz model. $N_x = 400$ and $\gamma = \rho = 0.1$.

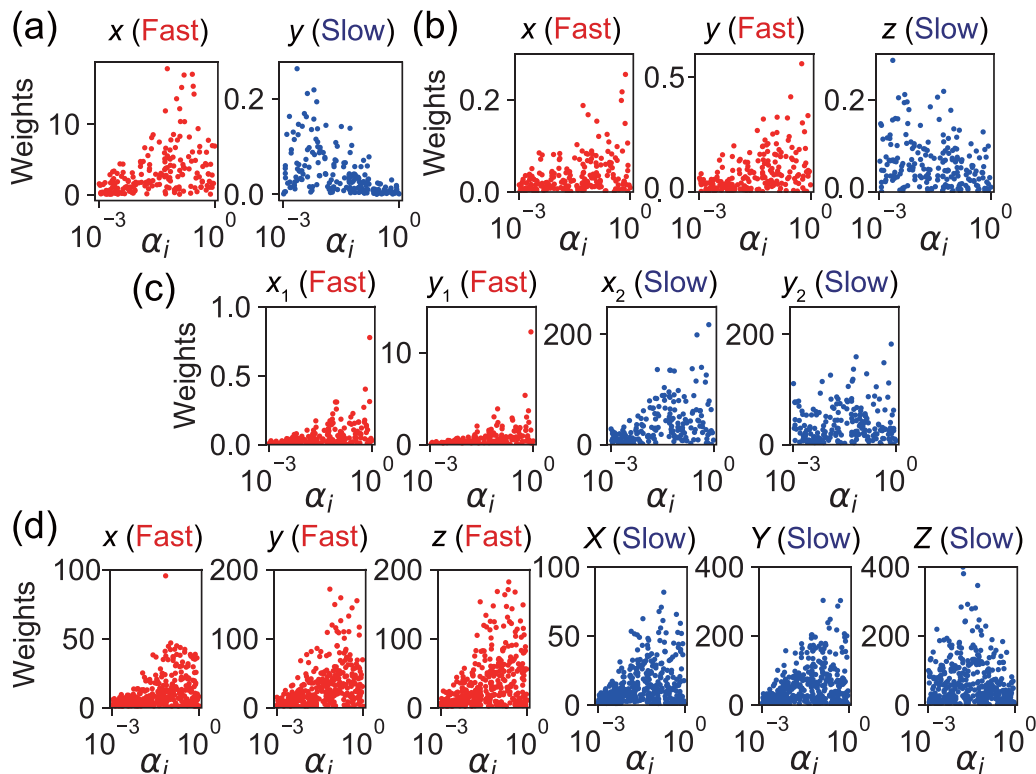


FIG. 4. The absolute output weights in \hat{W}^{out} of trained DTS-ESNs, plotted against the leak rates of the corresponding reservoir neurons. Each panel corresponds to each subsystem. The parameter values are the same as those for Fig. 3 but with $\alpha_{\min} = 10^{-3}$. (a) The Rulkov model. (b) The HR model. (c) The tc-VdP model. (d) The tc-Lorenz model.

Analyses. The timescales of the reservoir dynamics in the DTS-ESN are mainly determined by the hyperparameters including the time step Δt , the spectral radius ρ , and the leak rate matrix A . The timescales of reservoir dynamics in the linearized system of Eq. (1), denoted by τ_i for $i = 1, \dots, N_x$, are linked to the set of eigenvalues λ_i of its Jacobian matrix as follows (see the Supplemental Material [36]):

$$\tau_i = -\frac{\Delta t}{\ln |\lambda_i|}. \quad (5)$$

Figure 2 demonstrates that the timescale distribution in the linearized system is different between the LI-ESN and the DTS-ESN. Figure 2(a) shows a timescale distribution for the LI-ESN with $\alpha = 10^{-0.1}$. When α is decreased to produce slower dynamics, the mean of the timescale distribution increases while the distribution range in the logarithmic scale is almost unaffected as shown in Fig. 2(b) [18]. Figure 2(c) shows a broader timescale distribution of the DTS-ESN with $[\alpha_{\min}, \alpha_{\max}] = [10^{-3}, 1]$. As shown in Fig. 2(d), the distribution range monotonically increases with a decrease in α_{\min} (see the Supplemental Material [36]). The timescale analysis based on the linearized systems indicates that the DTS-ESN with a sufficiently small α_{\min} has much more diverse timescales than the LI-ESN.

Results. We evaluated the computational performance of the DTS-ESN in prediction tasks involved with four chaotic fast-slow dynamical systems models listed in Table I: (i) the Rulkov model which is a 2D map replicating chaotic bursts of neurons [32]; (ii) the Hindmarsh-Rose (HR) model which is a phenomenological neuron model exhibiting irregular spiking-

bursting behavior [33]; (iii) the two coupled Van der Pol (tc-VdP) model which is a combination of fast and slow limit cycle oscillators with nonlinear damping [34]; (iv) the two coupled Lorenz (tc-Lorenz) model which is a caricature representing the interaction between the ocean with slow dynamics and the atmosphere with fast dynamics [35,37]. There is a large timescale gap between the fast and slow subsystems.

We generated time series data from each dynamical system model using the parameter values listed in Table I. The ODE models were numerically integrated using the Runge-Kutta method with time step Δt . Then, we separated the whole time series data of total length T_{total} into transient data of length T_{trans} , training data of length T_{train} , and testing data of length T_{test} . The transient data was discarded to wash out the influence of the initial condition of the reservoir state vector. We set the time step and the durations of data as listed in Table II unless otherwise noted.

First, we performed the one-step-ahead prediction task (Task 1) using the open-loop model shown in Fig. 1(b), where the states of the whole variables at one step ahead are predicted only from the input time series of fast variables. The prediction performance is evaluated with the following normalized root mean squared error (NRMSE) between the model predictions and the target outputs:

$$\text{NRMSE} = \frac{\sqrt{\langle \|\mathbf{y}(t+k\Delta t) - \mathbf{d}(t+k\Delta t)\|_2^2 \rangle_k}}{\sqrt{\langle \|\mathbf{d}(t+k\Delta t) - \langle \mathbf{d}(t+k\Delta t) \rangle_k\|_2^2 \rangle_k}}, \quad (6)$$

where $\langle \cdot \rangle_k$ denotes an average over the testing period.

Figures 3(a)–3(d) show the comparisons between the DTS-ESN and the LI-ESN in the NRMSEs for the four dynamical systems models listed in Table I (see the Supplemental Material for examples of predicted time series [36]). The horizontal axis is α_{\min} for the DTS-ESN and α for the LI-ESN. When $\alpha_{\min} = \alpha = 1$, the two models coincide and yield the same performance. The prediction performance is improved as α_{\min} is decreased from 1 for the DTS-ESN, mainly due to an increase in the prediction accuracy with respect to the slow variables (see the Supplemental Material [36]). The DTS-ESN can keep a relatively small prediction error when α_{\min} is decreased even to 10^{-3} in contrast to the LI-ESN. The best performance of the DTS-ESN is obviously better than that of the LI-ESN for all the target dynamical systems, indicating a higher ability of the DTS-ESN. Moreover, even with α_{\min} fixed at 10^{-3} , the DTS-ESN can achieve the performance comparable to the best one obtained by the LI-ESN for all the target dynamics. This effortless setting of leak rates is an advantage of the DTS-ESN over the LI-ESN and the hierarchical LI-ESNs.

Figures 4(a)–4(d) demonstrate the absolute output weights in \hat{W}^{out} of the trained DTS-ESNs, plotted against the leak rates of the corresponding neurons. Each panel corresponds to an output neuron for the variable specified by the label on that. The results indicate that the reservoir neurons with large α_i , having small timescales, are mainly used for approximating the fast subsystems (red points) and those with small α_i , having large timescales, are for the slow subsystems (blue points). We can see that the neuronal states with appropriate timescales are selected to comply with the timescale of the desired output as a result of model training. In Fig. 4(d) for the tc-Lorenz model, the reservoir neurons with relatively small α_i are used for the slow variable Z but not for the other slow variables (X and Y). This means that the dynamics of X and Y are essentially not as slow as that of Z , causing the performance degradation with a large decrease in α_{\min} as shown in Fig. 3(d). By increasing the reservoir size and the length of training data, this degradation is mitigated (see Supplemental Material [36]). The natural separation of the roles of neurons can be regarded as a spontaneous emergence of modularization found in many biological systems [38–40].

Second, we performed the autoregressive prediction task (Task 2). As shown in Fig. 1(c), the open-loop model is trained using the training data, and then the closed-loop model is used to generate predicted time series autonomously. We increased T_{train} from 60 in Table II to 120 for the tc-Lorenz model, in order to improve the prediction performance. In the testing phase, we evaluated the valid time [6] for the slow dynamics, indicating the elapsed time duration (measured in actual time units) before the normalized prediction error $E(t)$ exceeds a threshold value ϵ in the testing phase. The error $E(t)$ is defined as follows [7]:

$$E(t) = \frac{\|\tilde{\mathbf{x}}(t) - \tilde{\mathbf{d}}(t)\|_2}{(\|\tilde{\mathbf{d}}(t + k\Delta t)\|_2^2)_k^{1/2}}, \quad (7)$$

where $\tilde{\mathbf{x}}(t)$ and $\tilde{\mathbf{d}}(t)$ represent the reservoir state vector and the target vector of slow variables, respectively, and $\langle \cdot \rangle_k$ denotes an average over the testing period.

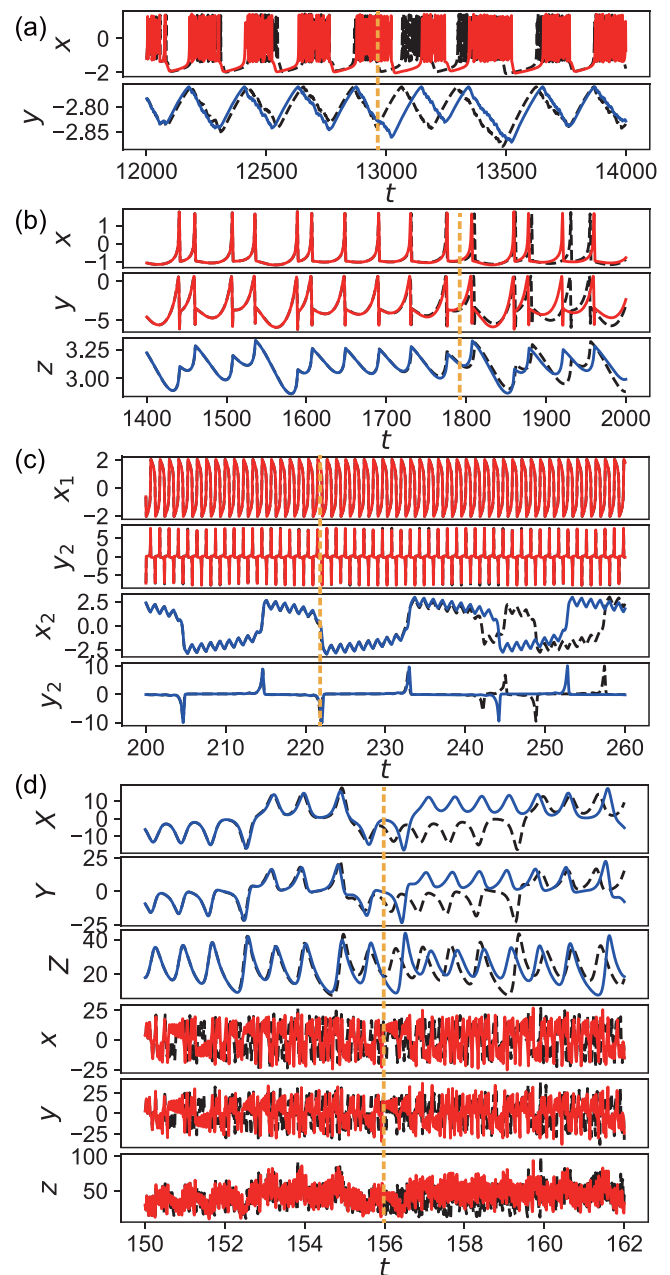


FIG. 5. Examples of predictions (red and blue solid lines) by the closed-loop DTS-ESN models, superimposed on the target time series (black dashed lines). The orange vertical dashed line indicates the valid time. (a) The Rulkov model. $N_x = 400$, $\gamma = 0.8$, $\rho = 1$, $\zeta = 1$, $\beta = 10^{-3}$, $\alpha_{\min} = 10^{-6/9}$, and $\epsilon = 0.01$. (b) The HR model. $N_x = 400$, $\gamma = 0.6$, $\rho = 0.2$, $\zeta = 0.4$, $\beta = 10^{-3}$, $\alpha_{\min} = 10^{-24/9}$, and $\epsilon = 0.05$. (c) The tc-VdP model. $N_x = 400$, $\gamma = 0.1$, $\rho = 0.03$, $\zeta = 0.2$, $\beta = 10^{-6}$, $\alpha_{\min} = 10^{-2/9}$, and $\epsilon = 0.4$. (d) The tc-Lorenz model. $N_x = 1000$, $\gamma = 0.01$, $\rho = 0.01$, $\zeta = 0.04$, $\beta = 10^{-4}$, $\alpha_{\min} = 10^{-8/9}$, and $\epsilon = 0.4$.

Figures 5(a)–5(d) show examples of autoregressive predictions (red lines for fast variables and blue lines for slow ones) by the closed-loop DTS-ESN models for the four target dynamical systems. The predicted time series approximates the target time series well (black dashed lines) until the valid time indicated by the vertical dashed line. The divergence

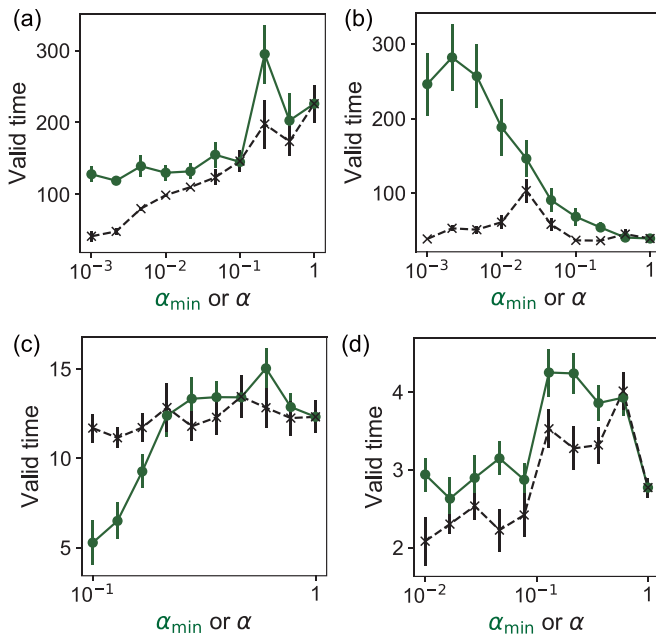


FIG. 6. Comparisons of the valid time between the DTS-ESN (green circles) and the LI-ESN (black crosses). The marks indicate the average values over 20 simulation runs with different reservoir realizations. The error bar indicates the standard error. The parameter conditions and data settings are the same as those for Fig. 5, except for the varied one. (a) The Rulkov model. (b) The Hindmarsh-Rose model. (c) The tc-VdP model. (d) The tc-Lorenz model.

of the prediction error after a finite time is inevitable due to the chaotic dynamics. We note that, in Fig. 5(c) for the tc-VdP model, the discrepancy between the predicted and target time courses is not prominent until around $t = 240$ but the normalized error exceeds $\epsilon = 0.4$ at around $t = 221$. If the threshold value is changed to $\epsilon = 3$, the valid time is increased to around 44. Figures 6(a)–6(d) demonstrate the comparisons of the valid time between the closed-loop DTS-ESN and the

closed-loop LI-ESN for the four target dynamical systems. In all the panels, the largest valid time is achieved by the DTS-ESN, suggesting its higher potential in the long-term prediction. The valid time largely depends on the hyperparameter setting including the range of leak rates, which may be associated with the attractor replication ability of the DTS-ESN as measured by the Lyapunov exponents (see Supplemental Material [36] for details).

Discussion and conclusion. We have proposed the RC model with diverse timescales, the DTS-ESN, by incorporating distributed leak rates into the reservoir neurons for modeling and prediction of multiscale dynamics. The results of the prediction tasks indicate the effectiveness of our randomization approach to realizing a reservoir with rich timescales. Although we assumed a specific distribution of leak rates in this study, another distribution could further improve the prediction performance. Moreover, another type of heterogeneity of reservoir components could boost the ability of RC systems in approximating a wide variety of dynamical systems. Future applications include a prediction of large-scale spatiotemporal chaotic systems [6,41] from partial observations and an inference of slow dynamics from experimentally measured data involved in real-world multiscale systems [31,42].

Acknowledgments. We thank D. Inoue for fruitful discussion and H. Ma for valuable comments. This work was performed as a project of Intelligent Mobility Society Design, Social Cooperation Program (Next Generation Artificial Intelligence Research Center, the University of Tokyo with Toyota Central R&D Labs., Inc.), and supported in part by JSPS KAKENHI Grants No. 20K11882 (G.T.), No. JP20H05921 (K.A.), JST-Mirai Program Grant No. JPMJMI19B1 (G.T.), JST CREST Grant No. JPMJCR19K2 (G.T.), JST Moonshot R&D Grant No. JPMJMS2021 (K.A.), AMED under Grant No. JP21dm0307009 (K.A.), and International Research Center for Neurointelligence at The University of Tokyo Institutes for Advanced Study (UTIAS) (G.T., K.A.).

- [1] D. G. Vlachos, A review of multiscale analysis: examples from systems biology, materials engineering, and other fluid–surface interacting systems, *Adv. Chem. Eng.* **30**, 1 (2005).
- [2] K. Matouš, M. G. Geers, V. G. Kouznetsova, and A. Gillman, A review of predictive nonlinear theories for multiscale modeling of heterogeneous materials, *J. Comput. Phys.* **330**, 192 (2017).
- [3] T. S. Deisboeck, Z. Wang, P. Macklin, and V. Cristini, Multiscale cancer modeling, *Annu. Rev. Biomed. Eng.* **13**, 127 (2011).
- [4] J. Walpole, J. A. Papin, and S. M. Peirce, Multiscale computational models of complex biological systems, *Annu. Rev. Biomed. Eng.* **15**, 137 (2013).
- [5] E. Weinan, *Principles of Multiscale Modeling* (Cambridge University Press, 2011).
- [6] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach, *Phys. Rev. Lett.* **120**, 024102 (2018).
- [7] J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. R. Hunt, M. Girvan, and E. Ott, Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model, *Chaos* **28**, 041101 (2018).
- [8] M. Alber, A. B. Tépole, W. R. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W. W. Lytton, P. Perdikaris, L. Petzold *et al.*, Integrating machine learning and multiscale modeling perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences, *npj Digital Medicine* **2**, 115 (2019).
- [9] H. Jaeger, The “echo state” approach to analysing and training recurrent neural networks, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148 (2001).
- [10] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, *Science* **304**, 78 (2004).
- [11] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural

- computation based on perturbations, *Neural Comput.* **14**, 2531 (2002).
- [12] D. Verstraeten, B. Schrauwen, M. D’Haene, and D. Stroobandt, An experimental unification of reservoir computing methods, *Neural Networks* **20**, 391 (2007).
- [13] M. Lukoševičius and H. Jaeger, Reservoir computing approaches to recurrent neural network training, *Comput. Sci. Rev.* **3**, 127 (2009).
- [14] P. J. Werbos, Backpropagation through time: what it does and how to do it, *Proc. IEEE* **78**, 1550 (1990).
- [15] M. Lukoševičius, A practical guide to applying echo state networks, in *Neural Networks: Tricks of the Trade* (Springer, 2012), pp. 659–686.
- [16] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert, Optimization and applications of echo state networks with leaky-integrator neurons, *Neural Networks* **20**, 335 (2007).
- [17] H. Jaeger, Discovering multiscale dynamical features with hierarchical echo state networks, Tech. Rep. (Jacobs University Bremen, 2007).
- [18] L. Manneschi, M. O. Ellis, G. Gigante, A. C. Lin, P. Del Giudice, and E. Vasilaki, Exploiting multiple timescales in hierarchical echo state networks, *Front. Appl. Math. Stat.* **6**, 616658 (2021).
- [19] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Recent advances in physical reservoir computing: A review, *Neural Networks* **115**, 100 (2019).
- [20] C. Gallicchio, A. Micheli, and L. Pedrelli, Deep reservoir computing: A critical experimental analysis, *Neurocomputing* **268**, 87 (2017).
- [21] H. Tamura and G. Tanaka, Transfer-RLS method and transfer-FORCE learning for simple and fast training of reservoir computing models, *Neural Networks* **143**, 550 (2021).
- [22] Z. Li and G. Tanaka, Multi-reservoir echo state networks with sequence resampling for nonlinear time-series prediction, *Neurocomputing* **467**, 115 (2022).
- [23] T. Akiyama and G. Tanaka, Computational efficiency of multi-step learning echo state networks for nonlinear time series prediction, *IEEE Access* **10**, 28535 (2022).
- [24] M. Di Volo and A. Destexhe, Optimal responsiveness and information flow in networks of heterogeneous neurons, *Sci. Rep.* **11**, 17611 (2021).
- [25] G. Tanaka, R. Nakane, T. Yamane, D. Nakano, S. Takeda, S. Nakagawa, and A. Hirose, Exploiting heterogeneous units for reservoir computing with simple architecture, in *International Conference on Neural Information Processing* (Springer, 2016), pp. 187–194.
- [26] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Information processing using a single dynamical node as complex system, *Nat. Commun.* **2**, 468 (2011).
- [27] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, Optoelectronic reservoir computing, *Sci. Rep.* **2**, 287 (2012).
- [28] F. Stelzer, A. Röhm, K. Lüdge, and S. Yanchuk, Performance boost of time-delay reservoir computing by non-resonant clock cycle, *Neural Networks* **124**, 158 (2020).
- [29] J. Pauwels, G. Verschaffelt, S. Massar, and G. Van der Sande, Distributed Kerr non-linearity in a coherent all-optical fiber-ring reservoir computer, *Front. Phys.* **7**, 138 (2019).
- [30] A. K. Seshadri, Fast–slow climate dynamics and peak global warming, *Climate Dynamics* **48**, 2235 (2017).
- [31] C. S. Meinen, R. C. Perez, S. Dong, A. R. Piola, and E. Campos, Observed ocean bottom temperature variability at four sites in the northwestern Argentine Basin: evidence of decadal deep/abyssal warming amidst hourly to interannual variability during 2009–2019, *Geophys. Res. Lett.* **47**, e2020GL089093 (2020).
- [32] N. F. Rulkov, Regularization of Synchronized Chaotic Bursts, *Phys. Rev. Lett.* **86**, 183 (2001).
- [33] J. L. Hindmarsh and R. Rose, A model of neuronal bursting using three coupled first order differential equations, *Proc. R. Soc. London B* **221**, 87 (1984).
- [34] K. P. Champion, S. L. Brunton, and J. N. Kutz, Discovery of nonlinear multiscale systems: Sampling strategies and embeddings, *SIAM J. Appl. Dynam. Syst.* **18**, 312 (2019).
- [35] G. Boffetta, P. Giuliani, G. Paladin, and A. Vulpiani, An extension of the Lyapunov analysis for the predictability problem, *J. Atmos. Sci.* **55**, 3409 (1998).
- [36] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.4.L032014> for further details.
- [37] F. Borra, A. Vulpiani, and M. Cencini, Effective models and predictability of chaotic multiscale systems via machine learning, *Phys. Rev. E* **102**, 052203 (2020).
- [38] N. Kashtan and U. Alon, Spontaneous evolution of modularity and network motifs, *Proc. Natl. Acad. Sci.* **102**, 13773 (2005).
- [39] D. M. Lorenz, A. Jeng, and M. W. Deem, The emergence of modularity in biological systems, *Phys. Life Rev.* **8**, 129 (2011).
- [40] G. R. Yang, M. R. Joglekar, H. F. Song, W. T. Newsome, and X.-J. Wang, Task representations in neural networks trained to perform many cognitive tasks, *Nat. Neurosci.* **22**, 297 (2019).
- [41] P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics, *Neural Networks* **126**, 191 (2020).
- [42] T. Proix, V. K. Jirsa, F. Bartolomei, M. Guye, and W. Truccolo, Predicting the spatiotemporal diversity of seizure propagation and termination in human focal epilepsy, *Nat. Commun.* **9**, 1088 (2018).