2022 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2022, Virtual, December 12-15, 2022

Spatiotemporal Chaotic Characteristics with Multidimensional Inputs in Echo State Networks

Takahiro Iinuma† and Sou Nobukawa†‡

†Graduate School of Information and Computer Science, Chiba Institute of Technology
2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan
‡Department of Preventive Intervention for Psychiatric Disorders
National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan
Email: s1731008FE@s.chibakoudai.jp (TI), nobukawa@cs.it-chiba.ac.jp (SN)

Abstract—An echo state network (ESN), consisting of an input layer, reservoir, and output layer, is a recurrent neural network (RNN) model with higher learning efficiency than other mainstream RNNs represented by gated RNNs. Recent various studies have attempted to input multidimensional data to an ESN, although the performance of ESN under multidimensional inputs deteriorates. To further proceed with this development, it is necessary to investigate the spatiotemporal response characteristics of ESNs under multidimensional inputs. We focused on the synchrony of states of reservoir neurons and investigated the ESN characteristics under multidimensional inputs using the measure of mean correlation coefficient of pairs among all reservoir neuron states and existing measures such as accuracy, memory performance, and maximum Lyapunov exponent (MLE). The results showed that in the case of low-dimensional inputs, the maximum memory performance is achieved as the MLE nears zero. In the case of high-dimensional inputs, maximum memory performance is achieved when the state exhibits less than zero MLE and minimum synchronization. An approach focusing on synchronization and the MLE contributes to the determination of the best ESN characteristics to achieve the maximized ESN performance under multidimensional inputs.

1. Introduction

Reservoir computing (RC) [1–3] is becoming a widely accepted approach for recurrent neural network (RNN) models with higher learning efficiency (i.e., fewer learning data and less learning time) than other mainstream RNNs represented by gated RNNs (i.e., long short-term memories (LSTMs) [4] and gated recurrent units (GRUs) [4]). Its high learning efficiency is achieved by two components of RC: a reservoir as the efficient feature extractor by spatiotemporal responses without learning and a readout as the simple transformer of the response into the desired output. Therefore, RC models have few learning parameters and use simple learning methods. Recent studies show that RC can be adopted to dynamic systems, which involve rapidly changing characteristics by virtue of high learning efficiency [3].

ORCID iDs Takahiro Iinuma: 0000-0001-8160-7434, Sou Nobukawa: 0000-0001-7003-6912

One typical RC model is the echo state network (ESN) [1, 2, 5, 6] (see Fig.1). It has an RNN with randomly fixed synaptic weights as the reservoir, which has spatial dynamics corresponding to the temporal input signal, and a linear transformer as the readout which can be learned by linear regression. This network architecture achieves high learning efficiency. However, the performance of an ESN is highly dependent on its hyperparameters, such as reservoir size (number of neurons in the reservoir) [2], spectral radius [2, 5, 6], and leaking rate [2]. Recent studies have shown appropriate ESN design methods [2, 7], while some aspects of them are empirical and still unclear and controversial. Carrol insisted that the edge of chaos for maximizing the ESN performance, which is conventionally defined as zero maximum Lyapunov exponent (MLE), must be redefined [8].

Various studies have attempted to input multidimensional data to an ESN, although its performance under multidimensional inputs deteriorates. Tong *et al.* used fixed-weight convolutional neural networks to encode four-dimensional time-series data as extracted low-dimensional features from images; these time-series data were subsequently input to an ESN [9]. Arrieta *et al.* transformed images into one-dimensional time-series data and input them into a four-layer deep ESN [10]. Subsequently, Arrieta *et al.* studied video input to ESNs, in which pixel data were directly input as multidimensional input to a four-layer deep ESN [10]. Additionally, the development of other ESN architectures, which maintain performance even under such inputs, is rapidly proceeding [11, 12].

To further proceed with this development, it is necessary to investigate the spatiotemporal response characteristics of ESNs under multidimensional inputs. In this context, this study's purpose is to reveal the character of the response of ESNs under multidimensional inputs by focusing on the synchrony of states of reservoir neurons. In this paper, we investigated the leaky integrator (LI) model, a type of ESN, and varied its spectral radius by using four measures: accuracy, memory performance, MLE, and correlation coef-

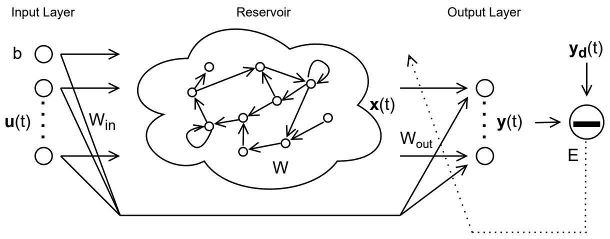

Figure 1: Overview of the echo state network (ESN) architecture. It consists of an input layer, a reservoir, and an output layer (readout).

ficients between neuron states as a measure of synchrony.

2. Model and Method

2.1. Echo State Network

2.1.1. ESN architecture

ESN architecture consists of three parts: an input layer, a reservoir, and an output layer (readout), as shown in Fig.1. In this paper, we used the LI model, which is an ESN with a mechanism to mix past states in leaking rate, α [6]. By that mechanism, the LI model can adjust to match the temporal scales between the input and reservoir dynamics. The neural dynamics update equation for the LI model reservoir when the time step is $t = 1, 2, \cdots$ is given by

$$\mathbf{x}(t+1) = (1-\alpha)\mathbf{x}(t) + \alpha \cdot f(W^{\text{in}}[\mathbf{u}(t); b] + W\mathbf{x}(t)). \quad (1)$$

The output is given by

$$\mathbf{y}(t) = W^{\text{out}}[\mathbf{x}(t); \mathbf{u}(t); b]. \quad (2)$$

Here, $[\cdot;\cdot]$ is a vector combination. $\mathbf{u}(t) \in \mathbb{R}^{N_u}$ is the input, $\mathbf{y}(t) \in \mathbb{R}^{N_y}$ is the output, and $\mathbf{x}(t) \in \mathbb{R}^{N_x}$ is the neuron state. b is the bias applied to the input, $f(\cdot)$ is the activation function, which is generally tanh, and α $(0 \leq \alpha \leq 1)$ is the leaking rate. The input weight $W^{\text{in}} \in \mathbb{R}^{N_x \times (N_u+1)}$ and the recurrence weight $W \in \mathbb{R}^{N_x \times N_x}$ are fixed with random weights. Exclusively, The output weight $W^{\text{out}} \in \mathbb{R}^{N_y \times (N_x+N_u+1)}$ is adjusted by learning.

2.1.2. Learning method

The ESN's learning method is ridge regression, a batch-learning method for the output weights of the readout. As the learning period is from t_s to t_e, the batch size is defined $T = t_e - t_s$. the output weights W^{out} is obtained by

$$W^{\text{out}} = ((X^T X + \beta I)^{-1} X^T Y_d)^T, \quad (3)$$

$$X = \{[\mathbf{x}(t_s); \mathbf{u}(t_s); b], \cdots, [\mathbf{x}(t_e); \mathbf{u}(t_e); b]\}^T,$$

$$Y_d = \{\mathbf{y}_d(t_s), \cdots, \mathbf{y}_d(t_e)\}^T,$$

where $\mathbf{y_d}(t) \in \mathbb{R}^{N_y}$ is the teacher signal, I is a identity matrix of size $N_x + N_u + 1$, β is the regularization factor, and $X \in \mathbb{R}^{T \times (N_x+N_u+1)}$, $Y_d \in \mathbb{R}^{T \times N_y}$.

2.1.3. Parameter settings

The two weights of the ESN, W^{in} and W, are determined by the following procedure. The input weight matrix, W^{in}, is determined with each element as a uniformly distributed random number in the range $[-\text{Scale}^{\text{in}}, \text{Scale}^{\text{in}}]$. However, the weight applied to the bias is in the positive range of $[0, \text{Scale}^{\text{in}}]$. The recurrence weight matrix, W, is determined by the following process and two parameters: the sparsity rate, s $(0 < s \leq 1)$, and the spectral radius, r $(0 < r)$. First, a uniform matrix with zero mean, $W_0 \in \mathbb{R}^{N_x \times N_x}$, is prepared. Next, W_0 is made sparse by replacing its elements with zeros at the sparsity rate, s. Finally, W is obtained from W_0 by

$$W = r\frac{W_0}{\rho(W_0)}, \quad (4)$$

where $\rho(W_0)$ is the maximum absolute value of eigenvalues of W_0.

2.2. Tasks of prediction time-series

The task used in the investigation is a time-series prediction task for random time-series delayed output (henceforth, the random delayed output task) [13]. The input signal at the t-time-step,$\mathbf{u}(t)$, is an N_u-dimensional vector of uniformly distributed random numbers in the range $[-1, 1]$, and the output signal is $\mathbf{y_d}(t) = \mathbf{u}(t-\tau)$. Here, τ is the delay between input and output signal; the larger τ required, the greater the memory capacity to solve, and the input and output dimensions are the same: $(N_u = N_y = \mathbb{N})$.

2.3. Evaluation method

We used four measures for our experiments: logarithmic normalized root mean squared error (log-NRMSE) [2] and memory capacity (MC) [13, 14] for ESN performance, MLE as the degree of chaotic behaviors [15], and correlation coefficient with delay between neurons as the degree of synchronization.

Table 1 lists the parameter settings for the experiment.

3. Results

We set experiment parameters as shown in Table 1, and investigated the log-NRMSE, MC, MLE, and correlation coefficient through changing the spectral radius under multidimensional inputs. Each value in the experimental results is the average of 10 samples. Figure 2 shows typical scatter plots among the MC, MLE, and correlation coefficient under one-dimensional inputs (corresponding to parts (a) and (b)) and 50-dimensional inputs (corresponding to parts (c) and (d)) in the case with delay $(\tau = 0)$ for correlation coefficient (shown in (a) and (c)) and the case with delay $(\tau = 20)$ (shown in (b) and (d)).

Under the one-dimensional inputs, the MC is maximized where the MLE is 0 (≈ -0.05) with a spectral radius near

Table 1: Parameter settings for the experiment.

Parameter	Value
Echo State Network	
input weight scale $\mathrm{Scale}^{\mathrm{in}}$	0.1
spectral radius r	0.05 to 2 in increments of 0.05
leaking rate α	1
sparsity s	0
reservoir size N_x	1000
regularization factor β	0.2
Task	
input/output dimension	$1, 2, 3, 4, 5, 10, 20, 30, 50, 75, 100, 125, 150, 200$
τ for log-NRMSE	$5, 20$
Evaluation method	
delay for correlation coefficient	$0, 20$

1 (0.97) (see (a) and (c) in Fig.2), which corresponds to findings of the conventional edge of chaos (MLE = 0). The point of the peak MC not only has a MLE close to 0, but also a small correlation coefficient.

Under 50-dimensional inputs, the memory capacity is maximized at a smaller spectral radius (0.92), i.e., where the MLE is less than 0 (≈ -0.2) and the correlation coefficient at delay = 20 is the lowest (see (d) in Fig.2). The correlation coefficient for delay = 0 shows a slight U-shape around the MLE 0 (see (c) in Fig.2). In contrast, the correlation coefficient for delay = 20 (see (d) in Fig.2) shows a relatively particular U-shape with the slight negative MLE (≈ -0.18) and the maximum peak of MC.

Similarly, we obtained results for log-NRMSE, shown as scatter plots with the same vertical and horizontal axes but the colors as log-NRMSE. The minimum log-NRMSE appeared to be independent of the correlation coefficient; however, it is dependent on the input dimension and task τ, and a larger spectral radius is optimal when they become larger.

4. Discussions and Conclusions

In this study, to reveal the spatiotemporal characteristics of the response of ESNs under multidimensional inputs, we gave ESNs tasks with low-dimensional to multidimensional inputs and investigated their accuracy (as log-NRMSE), memory performance (as MC), MLE, and synchronization (as correlation coefficients). The results showed that in the case of low-dimensional inputs, the maximum memory performance is at the MLE of zero. In the case of high-dimensional inputs, maximum memory performance is achieved at a slightly negative MLE < 0 and minimum synchronization. Moreover, the maximum memory performance clearly depends on the synchronization of reservoir neuron states. Conventionally, as a measure of maximizing ESN performance, the MLE of 0 has been considered as the edge of chaos [7], but Carroll showed the need for more complete measures to determine ESN performance [8]. Furthermore, asynchrony is known to play an important role in maintaining complex responses [16]. The high performances of ESN obtained in this study under multidimensional inputs are congruent with these findings. In conclusion, although more detailed evaluations under various kinds of multidimensional inputs are needed, the approach focusing on synchronization and MLE contributes to the determination of ESN characteristics to achieve the maximized ESN performance under multidimensional inputs.

References

[1] Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Héroux, 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.

[2] Mantas Lukoševičius. A practical guide to applying echo state networks. In *Neural networks: Tricks of the trade*, pp. 659–686. Springer, 2012.

[3] Kohei Nakajima and Ingo Fischer. Reservoir computing. *Springer*, Vol. 1, No. 5, p. 8, 2021.

[4] Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaee. Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*, 2017.

[5] Herbert Jaeger. 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, No. 34, p. 13, 2001.

[6] Herbert Jaeger, Mantas Lukoševičius, Dan Popovici, and Udo Siewert. Optimization and applications of echo state networks with leaky-integrator neurons. *Neural networks*, Vol. 20, No. 3, pp. 335–352, 2007.

[7] 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, 2012.

[8] Thomas L Carroll. Do reservoir computers work best at the edge of chaos? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 30, No. 12, p. 121109, 2020.

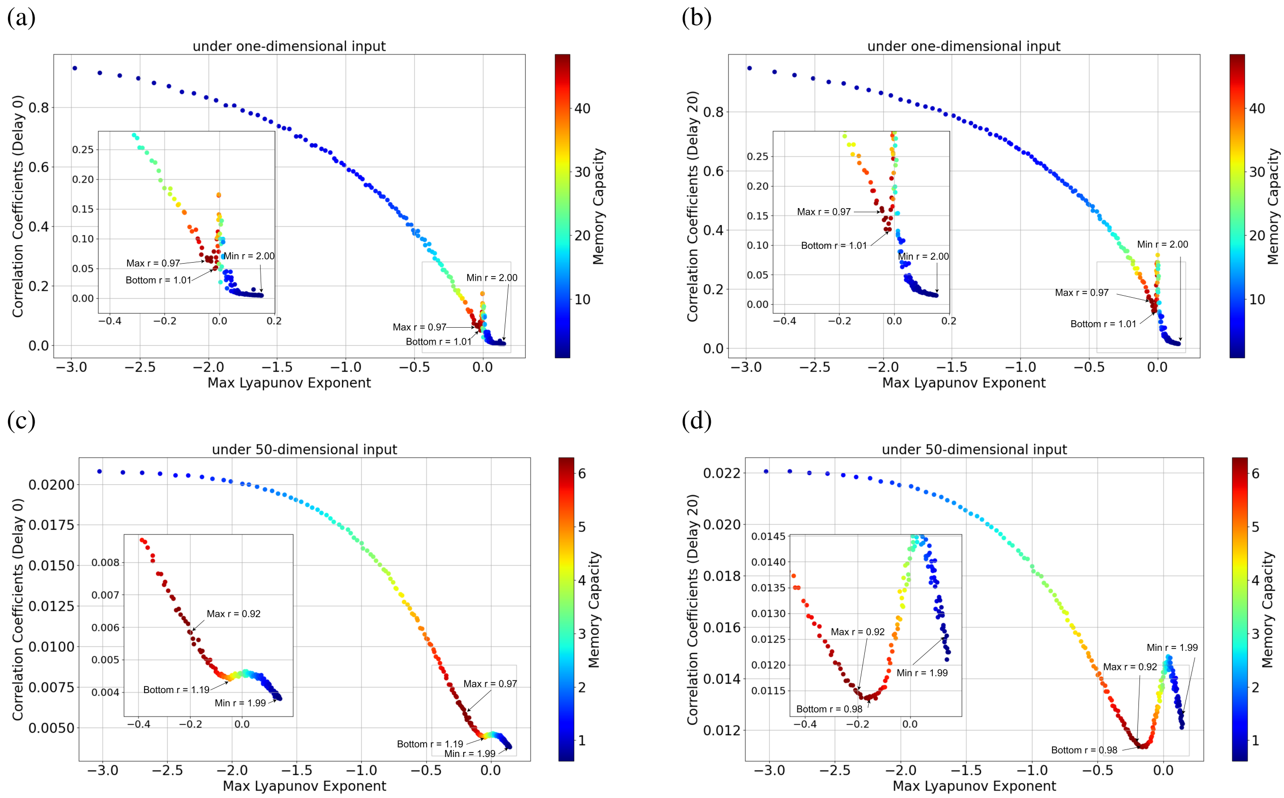

Figure 2: Scatter plots among memory capacity (as memory performance), MLE, and correlation coefficient (as synchronization) of the echo state network under one-dimensional inputs ((a) and (b)) and 50-dimensional inputs ((c) and (d)) in the case with delay ($\tau = 0$) for correlation coefficient ((a) and (c)) and the case with delay ($\tau = 20$) ((b) and (d)) (see Section 3). The vertical and horizontal axes show the correlation coefficient and MLE, respectively, and the colors represent memory capacity. The results show that in the case of one-dimensional inputs, the maximum memory capacity is at the MLE near 0. In the case of 50-dimensional inputs, the maximum memory capacity is at the MLE < 0 and minimum correlation coefficient. The "Bottom" note indicates the correlation coefficient, the "Max" and "Min" notes indicate the maximum and minimum MC, and the numbers following the notes are the corresponding spectral radius r.

[9] Zhiqiang Tong and Gouhei Tanaka. Reservoir computing with untrained convolutional neural networks for image recognition. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 1289–1294. IEEE, 2018.

[10] Alejandro Barredo Arrieta, Sergio Gil-Lopez, Ibai Laña, Miren Nekane Bilbao, and Javier Del Ser. On the post-hoc explainability of deep echo state networks for time series forecasting, image and video classification. *Neural Computing and Applications*, pp. 1–21, 2021.

[11] Matthew Evanusa, Snehesh Shrestha, Vaishnavi Patil, Cornelia Fermüller, Michelle Girvan, and Yiannis Aloimonos. Deep-readout random recurrent neural networks for real-world temporal data. *SN Computer Science*, Vol. 3, No. 3, pp. 1–12, 2022.

[12] Takahiro Iinuma, Sou Nobukawa, and Satoshi Yamaguchi. Assembly of echo state networks driven by segregated low dimensional signals. In *IEEE World Congress on Computational Intelligence (WCCI)*, p. #1814. IEEE, 2022.

[13] Herbert Jaeger. *Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the" echo state network" approach*, Vol. 5. GMD-Forschungszentrum Informationstechnik Bonn, 2002.

[14] Herbert Jaeger. Short term memory in echo state networks. gmd-report 152. In *GMD-German National Research Institute for Computer Science (2002), http://www. faculty. jacobs-university. de/hjaeger/pubs/STMEchoStatesTechRep. pdf*. Citeseer, 2002.

[15] Thomas S Parker and Leon Chua. *Practical numerical algorithms for chaotic systems*. Springer Science & Business Media, 2012.

[16] Sou Nobukawa and Haruhiko Nishimura. Chaotic resonance in coupled inferior olive neurons with the llinás approach neuron model. *Neural computation*, Vol. 28, No. 11, pp. 2505–2532, 2016.