entropy

MDPI

Article

Reservoir Dynamic Interpretability for Time Series Prediction: A Permutation Entropy View

Xiaochuan Sun [1,2], Mingxiang Hao [1,2], Yutong Wang [1,2], Yu Wang [1,2], Zhigang Li [1,2] and Yingqi Li [3,*]

[1] College of Artificial Intelligence, North China University of Science and Technology, Bohai Road, Tangshan 063210, China
[2] Hebei Key Laboratory of Industrial Perception, Tangshan 063210, China
[3] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China
* Correspondence: liyingqi@ncst.edu.cn

Abstract: An echo state network (ESN) is an efficient recurrent neural network (RNN) that is widely used in time series prediction tasks due to its simplicity and low training cost. However, the "black-box" nature of reservoirs hinders the development of ESN. Although a large number of studies have concentrated on reservoir interpretability, the perspective of reservoir modeling is relatively single, and the relationship between reservoir richness and reservoir projection capacity has not been effectively established. To tackle this problem, a novel reservoir interpretability framework based on permutation entropy (PE) theory is proposed in this paper. In structure, this framework consists of reservoir state extraction, PE modeling, and PE analysis. Based on these, the instantaneous reservoir states and neuronal time-varying states are extracted, which are followed by phase space reconstruction, sorting, and entropy calculation. Firstly, the obtained instantaneous state entropy (ISE) and global state entropy (GSE) can measure reservoir richness for interpreting good reservoir projection capacity. On the other hand, the multiscale complexity–entropy analysis of global and neuron-level reservoir states is performed to reveal more detailed dynamics. Finally, the relationships between ESN performance and reservoir dynamic are investigated via Pearson correlation, considering different prediction steps and time scales. Experimental evaluations on several benchmarks and real-world datasets demonstrate the effectiveness and superiority of the proposed reservoir interpretability framework.

Keywords: echo state network; time series prediction; interpretability; PE; reservoir richness; projection capability

check for updates

Citation: Sun, X.; Hao, M.; Wang, Y.; Wang, Y.; Li, Z.; Li, Y. Reservoir Dynamic Interpretability for Time Series Prediction: A Permutation Entropy View. *Entropy* **2022**, *24*, 1709. https://doi.org/10.3390/e24121709

Academic Editor: Ching-Hsue Cheng

Received: 5 September 2022
Accepted: 19 November 2022
Published: 23 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Reservoir computing (RC) [1] is widely recognized as a computational model suited for sequential data modeling. Its key component is the reservoir with a large number of sparsely and randomly connected neurons, capturing high-dimensional dynamic features of input data. Such an RC paradigm can avoid some drawbacks of gradient-descent RNN training, especially the time-consuming problem. The echo state network (ESN) is a popular RC model with low computational cost and powerful nonlinear projection capability [2]. Specially, the ESN training is fairly simple, since only the output weight is the training part, which is achieved by typical regression methods. Given these advantages, ESNs have been widely applied in the fields of time series prediction [3–5], image processing [6,7], feature extraction [8,9] and text classification [10]. Despite this fact, the ESN remains a black box due to its uninterpretable reservoir operation mechanism, i.e., high-dimensional projection [11,12]. It makes ESN hard to be a well-received paradigm for practical applications.

Currently, many efforts have been devoted to the RC interpretability for revealing its internal mechanism. Bianchi et al. pioneered the investigation of ESN interpretability

by analyzing reservoir neuron dynamics via recurrence plots (RP) [13]. Such RP could determine the stability of the network, and the metric analysis of RP helped adjust important hyperparameters to push the ESN into the stability boundary. Moreover, RP theory was also used to explain the self-organizing convolutional echo state network proposed by Lee et al. [14]. Bianchi et al. adopted a horizontal visibility graph approach to reflect ESN dynamics, which guided the adjustment of hyperparameters [15]. Ceni et al. suggested an excitable network attractor method to explain the operational mechanism of ESNs in specific tasks [16]. Variengien et al. proposed a recurrent state space visualization method, visualizing the learning process of ESNs, as well as revealing the effects of hyperparameters on reservoir dynamics [17]. Armentia et al. tried to illustrate how perturbed features affected the readout of ESNs using a perturbation-based importance attribution method [18]. Arrieta et al. presented a set of Explainable Artificial Intelligence techniques to visualize the potential memory, temporal patterns, and pixel absence effect of the model, thereby enabling the interpretation of DeepESN [19]. In the task of analyzing dynamic systems using ESN, Alao et al. interpreted the learned reservoir output weights as a representation of system dynamics through principal component analysis [20]. Baptista et al. used the SHapley Assitive exPlanations method for ESN to reveal the effect of different input features on the prediction results [21]. Some attempts devote to designing reservoir structures with high interpretability. Han et al. introduced an interpretable directed acyclic network for RC, where the effects of reservoir neurons on prediction performance were characterized by analyzing the memory property of each neuron [22]. Gauthier et al. designed an interpretable implicit RC model based on nonlinear vector autoregression to solve the reservoir uncertainty problem, i.e., random generation and multiple hyperparameters [1]. Miao et al. designed adaptive reservoirs based on the implementation theory of linear dynamical systems for a given task [23]. Despite these studies, for the reservoir with a large number of randomly sparsely connected neurons, it remains elusive how this uncertain structure can yield excellent performance. In other words, the correlation between reservoir richness and its high-dimensional projection capability remains unclear.

The permutation entropy (PE) [24] can reveal the complexity of time series by means of subsequence sorting in a high-dimensional reconstruction space [25]. The method is computationally simple, noise-resistant, and sensitive to local variations, thus possessing the powerful capacities of mutation information identification and visualization [26]. PE has been widely employed in fault detection [27], electrocardiography [28], complex systems [29], and other domains. To open the door for a wider acceptance of the ESN methodology, we explore the feasibility of the PE approach to explain good reservoir projection capability. The detailed motivation for the research in this paper can be found in the Section 2.1.

In this article, we tackle the interpretability issue of an ESN by analyzing the instantaneous and global dynamics of reservoir neurons with the PE measure. Such a method can quantify the information of the states in a high-dimensional dynamical systems over time. We believe that the PE analysis is a valuable tool for a deep insight of dynamic reservoir behavior. The contributions of this paper can be summarized as follows.

- We develop two reservoir dynamic analysis methods based on PE from the instantaneous and global modeling perspectives. This is the first attempt to use PE and its multiscale modeling tools to reveal the relationship between reservoir richness and projection capacity.
- We investigate the sensitivity of ISE and GSE on hyperparameters affecting reservoir richness.
- We use multiscale complexity–entropy to analyze the global reservoir and neuron-level states to verify the single-scale and input-driven properties of reservoirs.
- We reveal the multistep and multiscale relationships between ESN performance and reservoir dynamic, which is achieved by measuring the Pearson correlations between nonlinear approximation/memory capacity and global PE of neurons' states.

This paper is constructed as follows. In Section 2, we provide an introduction to the preparatory knowledge, which includes research motivation and ESN architecture. In Section 3, we provide the permutation entropy method and explain how reservoir dynamics can be investigated by employing PE. In Section 4, experiments are completed utilizing the PE approach, and the experimental data are examined for their insight into reservoir dynamics. In Section 5, we provide a detailed discussion on the relationship between ESN performance and reservoir state entropy. Finally, Section 6 presents final observations and future study directions.

2. Preliminary

2.1. Motivation

Since the ESN was proposed, a rich body of existing work has driven the progress of ESNs. Due to the ill-posed problem with the original ESN, much research has focused on topology construction, training methods, or hyperparameters tuning. However, the uninterpretable reservoir can greatly limit the application and advancement of ESN. Like traditional RNN, the reservoir states of ESN are continuously updated with time. Considering the temporal state of each neuron as a one-dimensional state, the whole reservoir state is regarded as a high-dimensional time-variant system. For this input-driven system, the reservoir layer neuron states are updated under the guidance of a randomly generated matrix W. Why is such a state update mode conducive to competitive reservoir projection and even high-efficient (remarkable) nonlinear approximation capacity? It is worth further study; thus, the focus of this paper is to explore the feasibility of revealing the good performance of ESN from the perspective of reservoir state analysis.

As stated in Section 1, this paper intends to focus on PE-based reservoir interpretability. In fact, Ozturk et al. defined the average state entropy (ASE) using Renyi's quadratic entropy and first proposed the concept of reservoir richness [30]. Furthermore, Gallicchio et al. used ASE as an indicator of reservoir complexity in DeepESN to support the applicability of multi-reservoir models [31]. However, both of them did not clarify the relationship between reservoir richness and reservoir projection capacity, and the significance of increasing reservoir richness is not discussed. On the other hand, ASE has been proposed only from the perspective of instantaneous reservoir states, neglecting more detailed neuronal time-varying states. Although Renyi's entropy can measure the sequence complexity, it is slightly inferior to PE in detecting dynamic mutations because it does not consider the ordering pattern of the system. Given these discussions, the motivation of this paper is to model the reservoir dynamics from multiple perspectives based on PE to reveal the correlation between reservoir richness and nonlinear projection capacity.

2.2. Esn Architecture

ESN is a popular RNN paradigm with a special hidden layer called a reservoir, which is generated from randomly and sparsely connected neurons. Such an information processing unit can effectively achieve the high-dimensional feature mapping of input data.

Figure 1 depicts the ESN structure. Obviously, ESN is composed of three layers: an input layer with K input units, a reservoir with N internal units, and an output layer with L output units. There exist the following types of connections, each of which has its own weight matrix, i.e., the input weight matrix $W^{in} \in R^{N\times K}$ for the weights from input units to internal ones, the internal weight matrix $W \in R^{N\times N}$ for the weights between internal units, and the output weight matrix $W^{out} \in R^{L\times(N+K)}$ for the weights from input and internal units to output ones. Generally, W^{in} and W are generated randomly and fixed during the ESN training. Only W^{out} (dashed line in Figure 1) is the trained part.

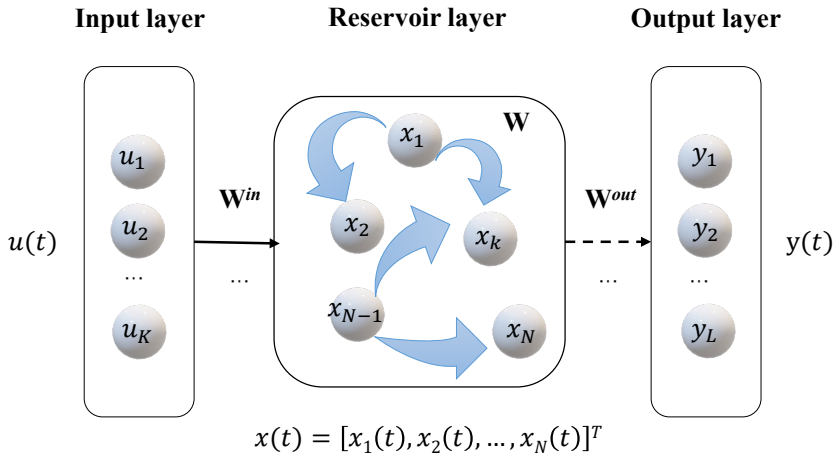


Figure 1. A sketch of the ESN architecture, where the output feedback of the reservoir is not considered.

For a given input signal $u(t) = [u_1(t), u_2(t), \ldots, u_K(t)]^T$, the state of the driven reservoir at time step t can be expressed as

$$x(t) = f(W^{in}u(t) + Wx(t-1)), \tag{1}$$

where $f(\cdot)$ is the activation function of internal units (sigmoid in our consideration). The corresponding network readout of this ESN is given by

$$y(t) = W^{out}x(t). \tag{2}$$

Once the internal state X and the desired output Y are collected, W^{out} can be effortlessly obtained by solving the following least squares problem

$$\min_{W^{out}} ||W^{out}X - Y||_2^2 \tag{3}$$

and its closed-form solution is calculated by a ridge regression in our scenario, that is

$$W^{out} = Y \cdot X^{-1} \tag{4}$$

It means the completion of the ESN training.

3. Methodologies

Here, we speak about how PE-based methodologies can be utilized to analyze the input-driven dynamics of an ESN reservoir. In Section 3.1, we discuss the PE algorithm in detail and describe our analysis method using the reservoir state as the analysis object; in Section 3.2, we discuss multiscale permutation entropy and statistical complexity measure (SCM) and how to use them to analyze the reservoir state of an ESN.

3.1. Permutation Entropy Measure of Reservoir Dynamics

Figure 2 gives an overview of our PE-enabled reservoir state analysis framework. In structure, this framework is composed of the following three functionalities, i.e., the extraction, PE modeling and PE analysis of reservoir states. First, two types of reservoir states are extracted during training, namely the instantaneous state of all reservoir neurons at certain moment and the individual neuron state over time. Then, the PE modeling is performed for entropy calculations of the two different reservoir state sequences extracted. Afterwards, the entropy analysis is used to quantitatively estimate the instantaneous and global reservoir entropies in order to measure the degree of reservoir richness.

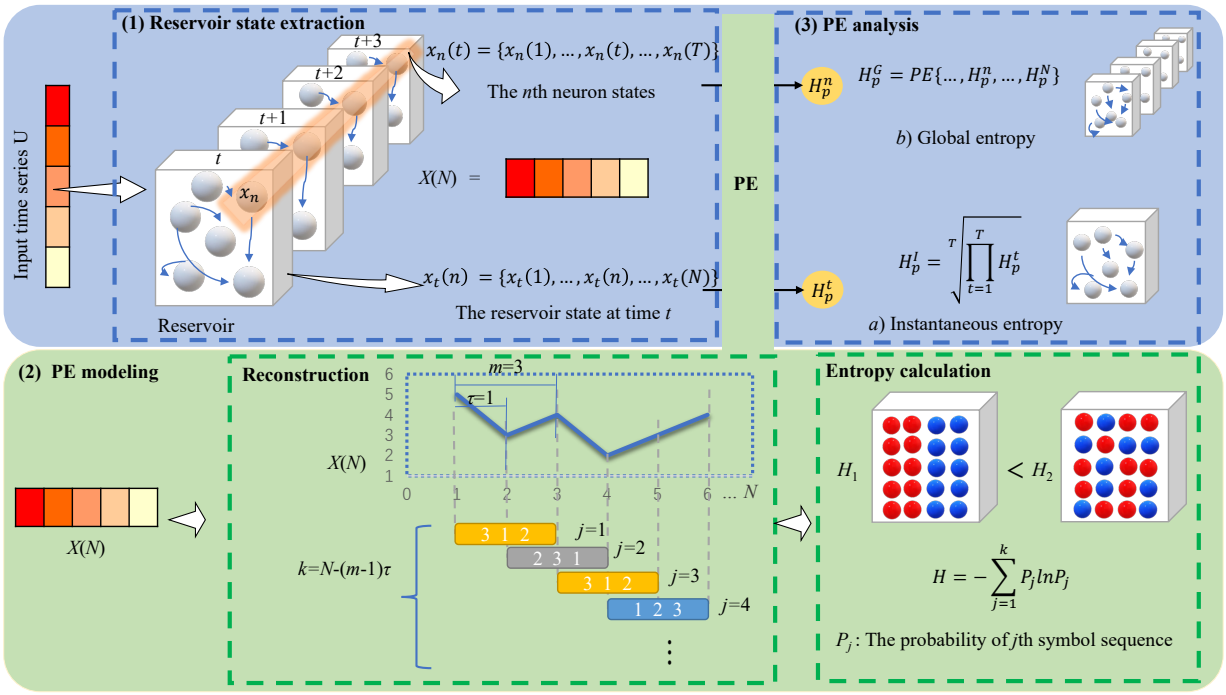


Figure 2. PE-based Interpretability framework for RC.

In the following, we elaborate the PE modeling of reservoir states, including phase space reconstruction, sorting, entropy calculation. Assuming the reservoir state $x_t(n) = \{x_t(1), \ldots, x_t(n), \ldots, x_t(N)\}$ at time t, we can obtain a reconstructed phase space with $k = N - (m-1)\tau$ components, which is described as follows

$$J = \begin{bmatrix} x_t(1) & x_t(1+\tau) & \ldots & x_t(1+(m-1)\tau) \\ \ldots & \ldots & & \ldots \\ x_t(j) & x_t(j+\tau) & \ldots & x_t(j+(m-1)\tau) \\ \ldots & \ldots & & \ldots \\ x_t(k) & x_t(k+\tau) & \ldots & x_t(k+(m-1)\tau) \end{bmatrix}, \tag{5}$$

where N is the number of reservoir neurons, while m and τ denote the embedding dimension and the delay time, respectively. Considering the jth reconstructed component $J(j)$, the m data in this component are rearranged in the order of smallest to largest. Specially, if any two elements of $J(j)$ are equal, their orders are not changed. Consequently, the sorted data can be expressed as follows

$$X_j = x_t(j + (l_1 - 1)\tau), x_t(j + (l_2 - 1)\tau), \cdots, x_t(j + (l_m - 1)\tau), \tag{6}$$

where $L = \{l_1, l_2, \ldots, l_m\}$ denote the column index of each element of the phase space after sorting. The m-dimensional phase space has the possible permutations up to $m!$. Using the Shannon entropy calculation method and normalizing, we can obtain the state entropy H_p^t of all neurons in the reservoir at time t. Furthermore, for the nth neuron with reservoir state $x_n(t) = \{x_n(1), \ldots, x_n(t), \ldots, x_n(T)\}$, the time-varying state entropy H_p^n of a given neuron can be found by the above method. The formula for calculating H_p^t and H_p^n is given as

$$\left\{H_p^t, H_p^n\right\} = -\frac{\sum_{j=1}^{k} P_j ln Pj}{ln(m!)}, \tag{7}$$

where $P_1, P_2, \ldots, P_k$ denote the occurring probability of each sequence. Obviously, the entropy calculations of these two different reservoir states follow the same expression.

Finally, our PE analysis focuses on the ISE and the GSE of the reservoir. Based on Equation (7), ISE can be obtained by the geometric mean of all H_p^t over all moments, which is denoted as follows

$$H_p^I = \sqrt[T]{\prod_{t=1}^{T} H_p^t}, \tag{8}$$

where T denotes the entire training time. The reservoir GSE is calculated by solving the normalized Shannon entropy of the sequence $\left\{H_p^1, \ldots, H_p^n, \ldots, H_p^N\right\}$ consisting of the time-varying state entropies of all reservoir neurons, which are given by

$$H_p^G = PE\{H_p^1, \ldots, H_p^n, \ldots, H_p^N\}. \tag{9}$$

In our consideration, ISE aims at reflecting the nonlinear projection capability of the reservoir by the measure of instantaneous reservoir richness. GSE considers the neuron level on the basis of ISE and reveals different factors affecting reservoir richness. The detail of these two reservoir dynamic analysis can be found in Algorithm 1.

Algorithm 1 PE analysis on reservoir states

Input: Datasets
Output: H_p^I, H_p^G

Initialize ESN
for $n \to N$ **do**
  for $t \to T$ **do**
    Collect reservoir states
  end for
end for
Obtain the ESN state matrix X
Extract the reservoir state: $x_t(n)$, $x_n(t)$
Initialize PE parameters m, τ
for $x_t(n)$, $x_n(t)$ **do**
  Obtain J_{km} based on Equation (5)
  for $j \to k$ **do**
    Obtain X_j based on sorting to $J(j)$
    Obtain the sequence of symbols L
  end for
  Calculate the probability P of occurrence of L
  Calculate H_p^t and H_p^n based on Equation (7)
end for
Obtain ISE H_p^I and GSE H_p^G based on Equations (8) and (9)
return H_p^I, H_p^G

3.2. Multiscale Complexity-Entropy Analysis

To further explore the reservoir dynamics, here, we use a multiscale complex permutation entropy (MCPE) method. The multiscale permutation entropy and statistical complexity are described below.

Multiscale permutation entropy (MPE) is an improvement PE. The basic idea is to first multiscale coarse-grain the time series and then calculate its PE. For a reservoir state $x_t = \{x_t(1), \ldots, x_t(n), \ldots, x_t(N)\}$ at time t, the versions x_j^s at different scales can be obtained by the following coarse-grained decomposition:

$$x_j^s = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x_i, \tag{10}$$

where s is the scale factor, and j is the sequence index after coarse granulation, following $1 \leq j \leq \frac{N}{s}$. Specially, when $s = 1$, the MPE is the traditional PE. After the above decomposition, using Equations (7) and (10), we can obtain the GSE H_p^G and the neuron-level H_p^n at different scales, respectively, i.e., the MPE.

Furthermore, we give the SCM of reservoir dynamics, which is derived from the product of the entropy and the disequilibrium. This measure is able to grasp the essential details of dynamics while discerning different degrees of periodicity and chaos. Given a probabiliby distribution P in Equation (7), the SCM can be defined by the product of the normalized permutation entropy H_p and a suitable metric distance $Q_j[P, P_e]$, which is expressed as follows:

$$C_{js} = Q_j[P, P_e]H_p. \tag{11}$$

In this formula, H_p refers to the H_p^G in GSE perspective, or the H_p^n at the neuron-level, therefore, C_{js} stands for the global reservoir complexity or single neuron complexity, respectively, and $Q_j[P, P_e]$ represents the degree of difference of the probability $P = \{P_1, \ldots, P_k\}$ from the uniform distribution $P_e = \{P_{e_1}, \ldots, P_{e_k}\} = \{\frac{1}{m!}, \ldots, \frac{1}{m!}\}$, which is given by

$$Q_j[P, P_e] = \frac{S[(P + P_e)/2] - S[P]/2 - S[P_e]/2}{Q_{max}}, \tag{12}$$

where S denotes the un-normalized information entropy, and Q_{max} is the maximum possible value of $Q_j[P, P_e]$, which is calculated as follows:

$$Q_{max} = -\frac{1}{2}[\frac{m! + 1}{m!}\log(m! + 1) - 2\log(2m!) + \log(m!)], \tag{13}$$

where $m!$ is all possible permutations mentioned in Section 3.1.

The combination of the above MPE and SCM measures form our MCPE approach for interpreting reservoir dynamics. On the one hand, the MCPE is able to demonstrate time series complexity trends at multiple scales. In our consideration, it is used to detect reservoir multi-scale dynamics [32]. On the other hand, it can distinguish random and chaotic behaviors through delineating the representation space known as the complexity entropy causal plane (CECP) [33].

4. Experiments

In this section, the corresponding experiment evaluation is conducted to verify the superiority of the proposed interpretability method. We consider four different datasets, including three classical benchmark tasks for time series modeling and a real-world dataset. First, we devote to revealing its remarkable nonlinear projection from the views of instantaneous and global reservoir dynamic in the PE analysis framework. It can be achieved by the impact analysis of reservoir hyperparameters on its state entropy experimentally. On the other hand, such reservoir dynamics is further dissected more deeply by a multi-scale approach.

In the following experiments, we consider an ESN with no output feedback, where W^{in} and W are randomly generated in the interval $[-1, 1]$, the connectivity of W is 0.05, and W^{out} is trained by the ridge regression with the regularization factor of 10^{-6}. To highlight the performance comparison, we add the noise obeying a Gaussian distribution with the variance of 0.1 for the used four datasets. During ESN training, according to the standard dropout procedure, the first 100 elements of the training data are discarded to remove the transients of the ESN. Specially, we use a fixed invariant random number seed in all experiments, and the permutation entropy is computed by the ordpy Python library proposed by Pessa et al. [34].

4.1. Dataset

Three benchmark time series and one real-world dataset are considered, including the chaotic mapping system Mackey–Glass (MG) and Lorenz, the stochastic model nonlinear autoregressive moving average (NARMA), and the New York crude oil market average price (CO). Figure 3 shows the trends of these time series over time.

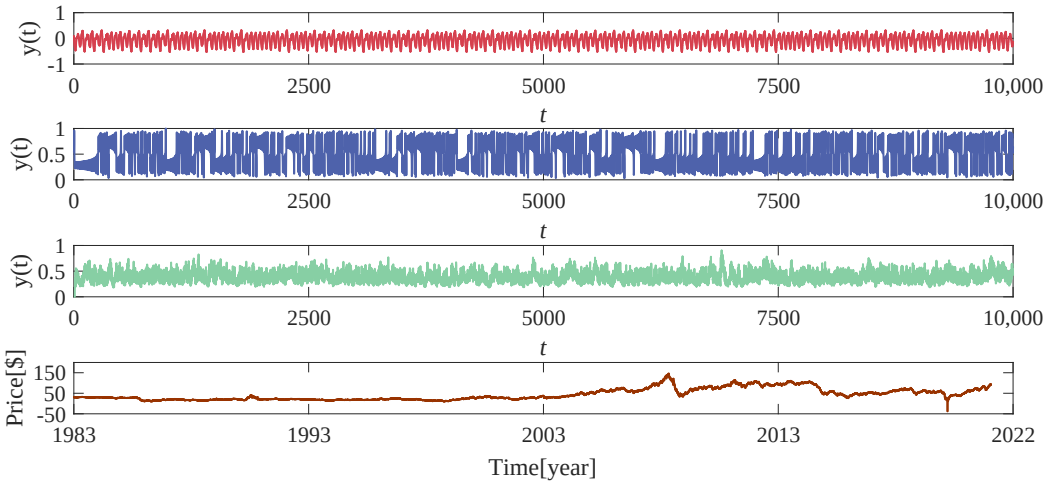


Figure 3. A sample of data. From top to bottom are the MG system, the Lorenz system, the NARMA system, and the CO for real-time timing data. The horizontal coordinate indicates the time step, a total of 10,000.

As is known, ESN is preferred in dealing with chaotic time series by virtue of its outstanding nonlinear projection capability on MG and Lorenz systems. The MG system is generated by the following equation:

$$\frac{dy(t)}{dt} = \frac{0.2y(t-\tau)}{1+y^{10}(t-\tau)} - 0.1y(t), \tag{14}$$

where the whole sequence is chaotic, acyclic, and non-divergent (convergent) when $\tau > 17$, and 10,000 sample points are used at $\tau = 17$. The Lorenz chaotic attractor [35] is defined as follows

$$\begin{aligned} \dot{x} &= \sigma(y-x) \\ \dot{y} &= x(\rho-z)-y \\ \dot{z} &= xy-\beta z+x, \end{aligned} \tag{15}$$

In this formula, $\sigma = 10$, $\beta = 8/3$, $\rho = 28$ in general.

As a typical discrete-time system, NARMA has been widely utilized for the performance assessment of neural networks. Effectively modeling NARMA is a difficult task due to its nonlinearity and long-memory problems. In our experiments, the tenth-order NARMA system is considered, which is generated by

$$y(t+1) = 0.3y(t) + 0.05y(t)[\sum_{i=0}^{9} y(t-i)] + 1.5u(t-9)u(t) + 0.1, \tag{16}$$

where $u(t)$ and $y(t)$ denote the input and output at moment t, respectively, which are randomly yielded from a uniform distribution with the interval $[0, 1]$.

Finally, the real-world dataset on the daily closing spot price of the West Texas Intermediate is used for our experiment evaluation. It can be collected from the US Energy Information Administration website over the time range from 4 April 1983 to 23 February 2022 ($N = 9766$ oil price observations), which is quoted in US dollars per barrel. As proved

in [36], oil price time series have strong stochastic properties; thus, its accurate prediction is exceedingly challenging.

4.2. ISE Analysis

In Figure 4, we plot the state entropy of all reservoir neurons at each moment for different time series prediction tasks, i.e., H_p^t. From this figure, it is obvious that these normalized entropy values fluctuate in the range between 0.995 and 1. Such a high PE over all moments reflects the considerable complexity of different neuron states, indicating significant differences among these reservoir neurons [37]. It can effectively alleviate the collinearity issue of the reservoir, thus obtaining an excellent nonlinear projection capability [38]. On the other hand, the state entropy of the reservoir always fluctuates steadily within a specific range, which means that the reservoir retains high richness over time, even though each neuron's state continuously updates. In addition, the entropy of reservoir states is task dependent, interpreting the input-driven mechanism of ESNs.

To explicitly illustrate the effects of hyperparameter tuning on reservoir richness, we visualize the ISE variation against spectral radius ρ and reservoir size N applied to the MG, lorenz, NARMA and CO datasets, as shown in Figure 5. From this figure, there exists the significant fluctuation of H_p^I as N decreases, meaning that the high entropy in low-size reservoirs hardly occurs. For the small N, the reservoir dynamic appears to be more sensitive to spectral radius. Intuitively, $\rho \in [0.6, 1]$ seems to be the best choice in the modeling of NARMA and CO datasets, while the relatively small or large ρ should be considered for MG and Lorenz, respectively, i.e., $\rho \in [0.2, 0.7]$ or $\rho > 1$. In the contrary, the large-scale reservoir $(N \geq 150)$ is more conducive to yielding high ISE, and in this case, ρ has a slight effect on the ISE. Such a finding can be illustrated by the fact that a large reservoir size can determine the richness of a randomly connected reservoir. While the spectral radius is more suitable for ensuring the echo state and stable edges of the ESN, the effect on richness is uncertain [39].

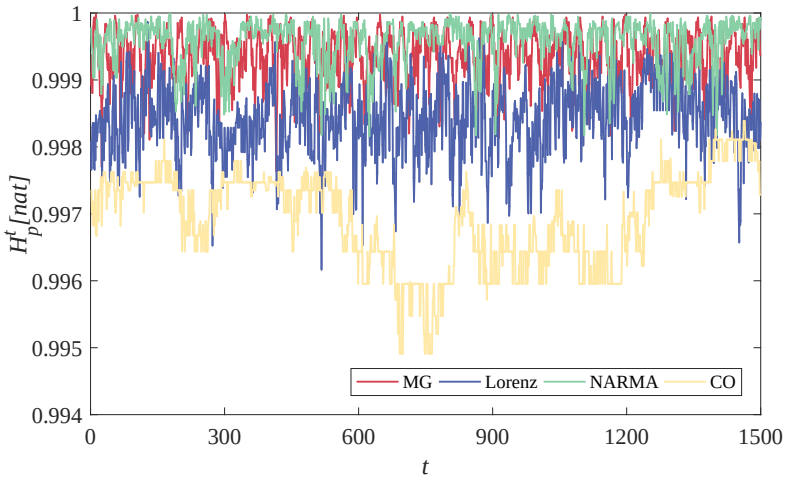


Figure 4. PE of the normalized reservoir states over time for the considered four time series prediction tasks, where the horizontal coordinate denotes the whole state updating steps, and the vertical coordinate denotes the corresponding PEs.

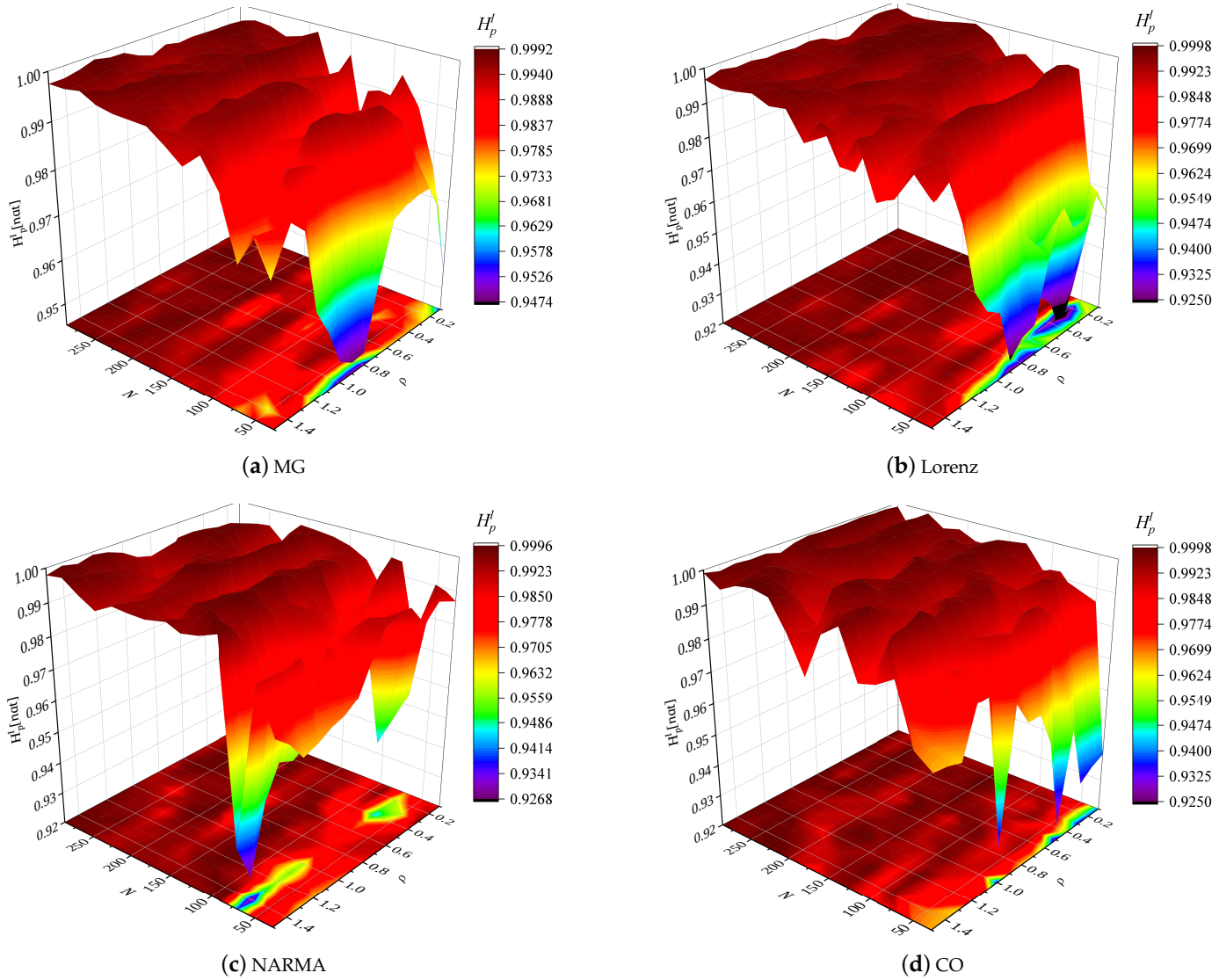


Figure 5. Reservoir ISEs versus reservoir size N and spectral radius ρ in the MG, Lorenz, NARMA and CO prediction tasks.

4.3. GSE Analysis

As is known, the states of reservoir neurons are continuously updated during the training process. The data are gradually uploaded into the reservoir as time flows, and each neuron thus obtains a time-varying state sequence. Investigating the neuron-level time-varying effect can provide an insight of neuron-level projection. Figure 6 depicts the time-varying state entropy of each reservoir neuron for different time series prediction tasks, i.e., H_p^n. Obviously, from a global view, such neuron-level entropy is task dependent, where it fluctuates around consistently high levels for the CO and NARMA series rather than the MG series.

On the one hand, from the task view, the difference in the low and high entropy values is mainly caused by the different complexity of the input series, where the complexity is measured by the PE method, as listed in Table 1, and the bigger PE means the more significant complexity. On the other hand, for the same prediction task, the neurons generated from the randomly generated and sparsely connected reservoir have significant differences of entropy, reflecting different activation levels of the reservoir neurons [40]. This interindividual variability between neurons provides a support for our investigation of reservoir richness.

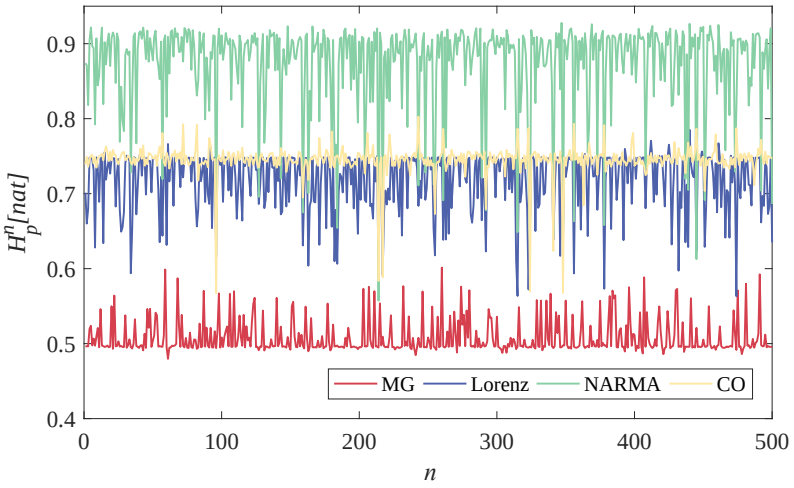
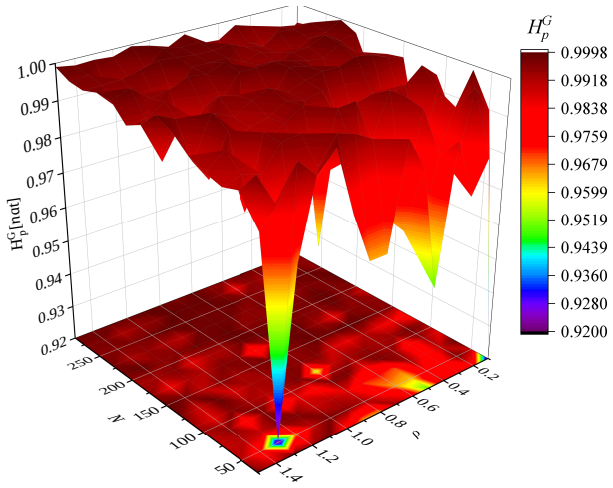


Figure 6. PEs of time-varying states of different reservoir neurons for the considered four time series prediction tasks, where the horizontal coordinate denotes each neuron in the reservoir, and the vertical coordinate denotes the corresponding PEs.

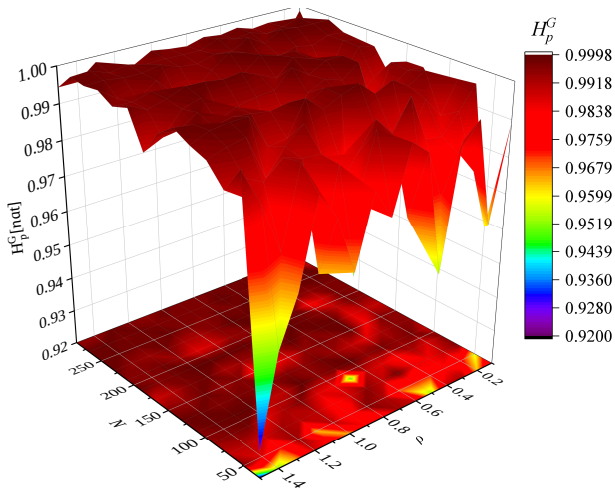
Table 1. Complexity measure on time series MG, Lorenz, NARMA and CO.

Indicator	MG	Lorenz	NARMA	CO
PE	0.55409	0.75621	0.97349	0.96773

Figure 7 evaluates the effects of N and ρ on the reservoir GSEs in different time series prediction tasks, where this global entropy is obtained by calculating the entropy of the time-varying state entropies of all reservoir neurons in Figure 6. Apparently, similar to the ISE visualization, the joint tuning of these two hyperparameters can enable the ESN to obtain an optimal GSE for these prediction tasks, guaranteeing reservoir richness, and the reservoir size is still the most influential hyperparameter for reservoir dynamic. Furthermore, we also evaluate the effects of input scaling ω_i and spectral radius ρ on the GSE in the case of $N = 200$, as shown in Figure 8. Obviously, the reservoir GSE always fluctuates at the level of high values, and ρ has a more significant effect on the GSE than ω_i, determining the reservoir projection capability more sensitively.

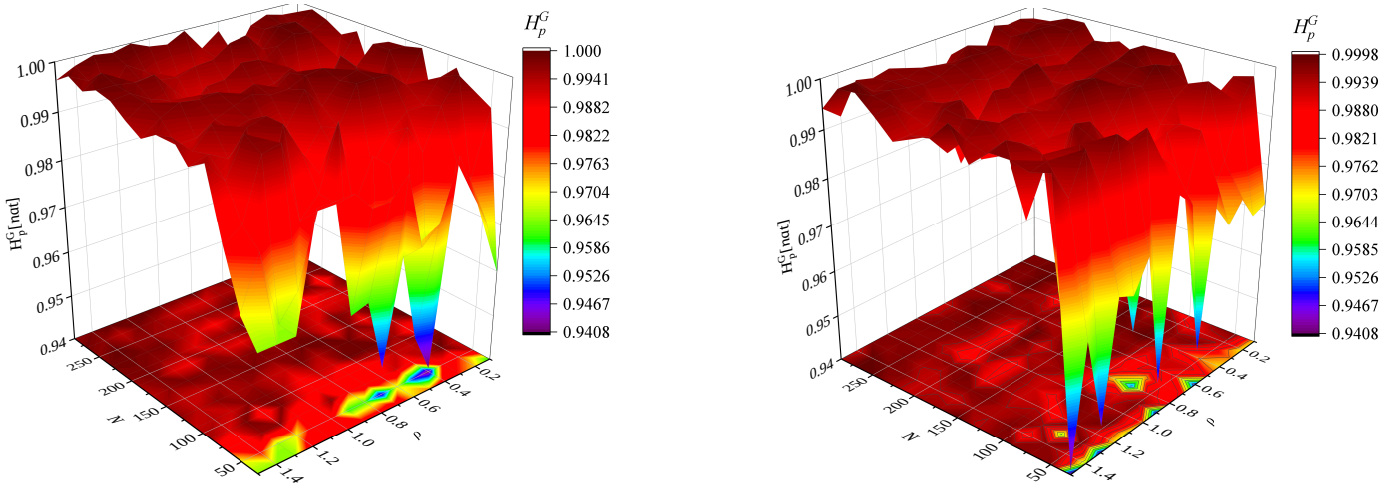


(**a**) MG



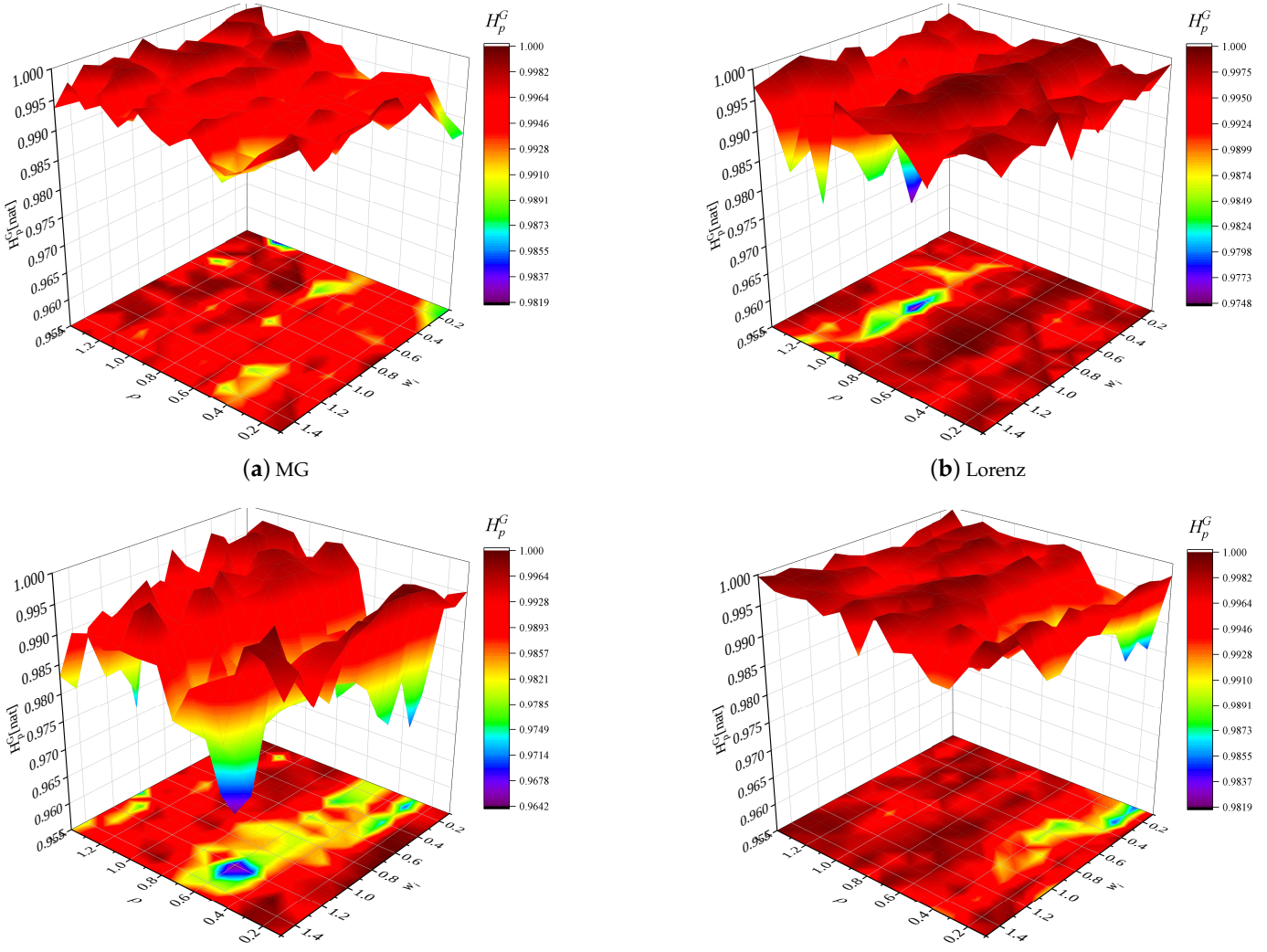
(**b**) Lorenz

Figure 7. *Cont.*



(**c**) NARMA (**d**) CO

Figure 7. Reservoir GSEs versus reservoir size N and spectral radius ρ in the MG, Lorenz, NARMA and CO prediction tasks.



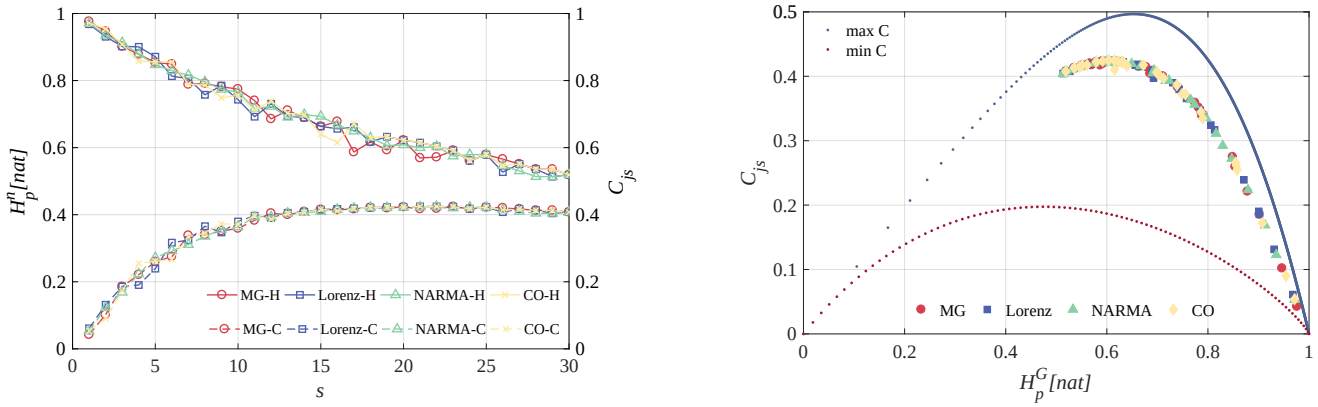
(**a**) MG (**b**) Lorenz

(**c**) NARMA (**d**) CO

Figure 8. Reservoir GSEs versus reservoir size N and spectral radius ρ in the MG, Lorenz, NARMA and CO prediction tasks.

4.4. Multiscale Reservoir Dynamics

Here, we concentrate on the MCPE experiments for global reservoir dynamic and its neuron-level dynamic, respectively, allowing the dissection of complex systems at multiple scales and helping characterize reservoirs more effectively. Figure 9 shows the GSE of reservoirs at different scales for the considered time-series prediction tasks as well as the relationship between complexity and GSE, i.e., the complexity–entropy causality plane (CECP). From the MCPE plot, it is observed that as the scale s increases, the GSE decreases continuously, while the statistical complexity C_{js} increases, and consequently, their changes tend to level off. It is noticeable that from our coarse-grained treatment, the reservoir has the highest GSE at the single scale, i.e., $s = 1$, and the larger s can yield a less complex reservoir. Such a single-scale phenomenon illustrates the significant difference between reservoir neuron states, meaning weak collinearity. Hence, the reservoir has a strong nonlinear projection capability. In Figure 9b, larger entropy values correspond to smaller C_{js}, which is a piece of strong evidence for reservoir stochasticity. The effect of different prediction tasks on reservoir richness is less obvious here.

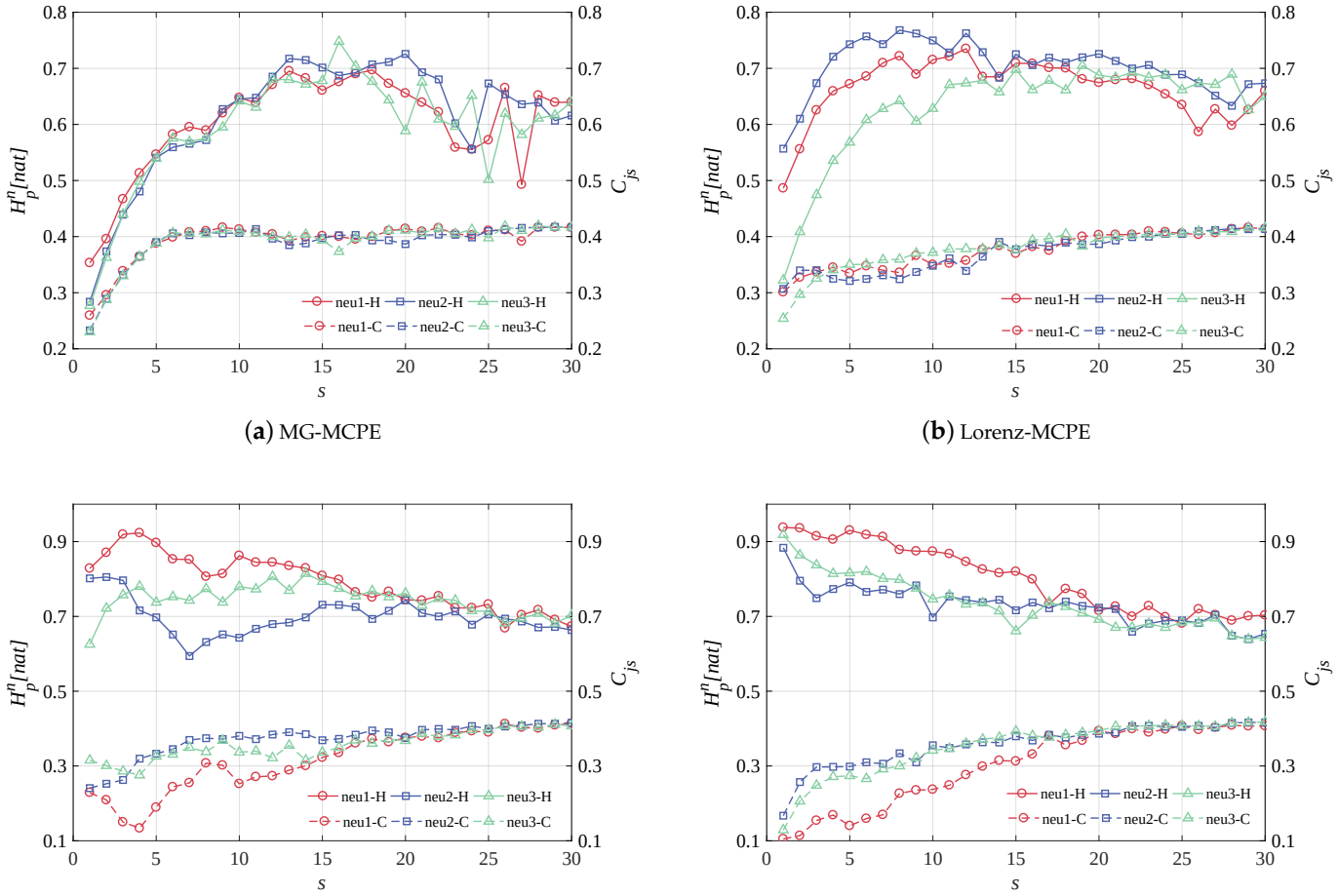


(**a**) MCPE (**b**) CECP

Figure 9. Multiscale global reservoir dynamics in the MG, Lorenz, NARMA and CO prediction tasks. (**a**) MCPE, where the solid and dashed lines are the variation curves of entropy and complexity, respectively, (**b**) CECP, where the top and bottom scatters represent the maximum and minimum complexity.

From a neuron-level view, Figure 10 depicts the multiscale dynamic metric of the time-varying state of a single reservoir neuron for different time series prediction tasks. For four data-driven reservoirs, we consider the same three neurons to reveal their time-varying state features. For the MG prediction task in Figure 10a, as s increases, H_p^n increases sharply and then decreases slowly, and C_{js} increases and then tends to be stable. The H_p^n of Lorenz-driven reservoir neurons increased rapidly at low scales, peaked at $s = 7$ and then leveled off. Compared with Figure 10a, the neuronal entropy values under the Lorenz system appear significantly different at low scales. For the NARMA and CO tasks, as s increases, H_p^n decreases and C_{js} slowly increases, as shown in Figure 10c,d. It is worth noting that the evident differences between neurons also appear at low scales. The above phenomenon demonstrates the difference in the activation level of neurons under different tasks. Because the time-varying states of neurons are driven by the inputs, the reservoir can effectively capture the input features, which ensures the ability of the reservoir to project nonlinearly for different tasks. In Figure 11, for the reservoir neuron of the MG task, the H_p^n only reaches a maximum of 0.7 with increasing C_{js}, which means that the neuron exhibits the same chaotic nature as the MG dataset. A similar situation also occurs for the Lorenz task, since the yielded H_p^n and C_{js} are scattered in the central region of the CECP. In addition, the other two neuron time-varying state sequences have the same stochastic nature as the input data NARMA and CO. The reservoir neurons have the same properties

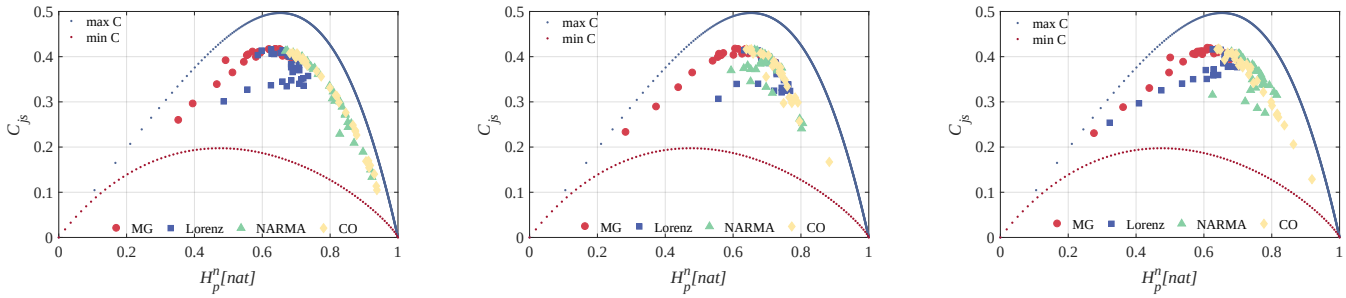
as the original input data, further supporting the preservation of the input features on the reservoir and revealing the stability of the reservoir.



(**a**) MG-MCPE (**b**) Lorenz-MCPE

(**c**) NARMA-MCPE (**d**) CO-MCPE

Figure 10. The MCPEs of three randomly selected neurons for the MG, Lorenz, NARMA, and CO datasets, where the solid line indicates the entropy change curve of H_p^n, and the dashed line indicates the complexity change curve of C_{js}.



(**a**) neu1-CECP (**b**) neu2-CECP (**c**) neu3-CECP

Figure 11. The CECPs of three randomly selected neurons for the MG, Lorenz, NARMA, and CO datasets, where the top and bottom lines represent the maximum and minimum complexity, respectively, and the different symbolic markers are used to distinguish the datasets.

5. Discussion

In this section, to highlight the critical contributions, we give a detailed discussion on the relationships between the ESN performance and reservoir dynamic. The former refers to nonlinear approximation performance and memory capacity (MC), while the latter

denotes the ISE and GSE of reservoirs. Specially, mean square error (MSE) is used for the approximation measure, which is given by

$$MSE = \frac{1}{l_{test}} \sum_{t=1}^{l_{test}} (y_{test}(t) - \hat{y}_{test}(t))^2, \tag{17}$$

where l_{test} is the test sample length, and $y_{test}(t)$ and $\hat{y}_{test}(t)$ are the test output and desired output at time t. The MC of ESN is defined as a reconstruction ability of the input from the past moment t, which is calculated as follows

$$MC = \sum_{k=1}^{\infty} MC_k. \tag{18}$$

where k denotes the input delay ($k = 1, 2, \ldots, 40$ in our scenario), and MC_k is actually the correlation coefficient between the input and the actual output by delaying k steps, which is given by

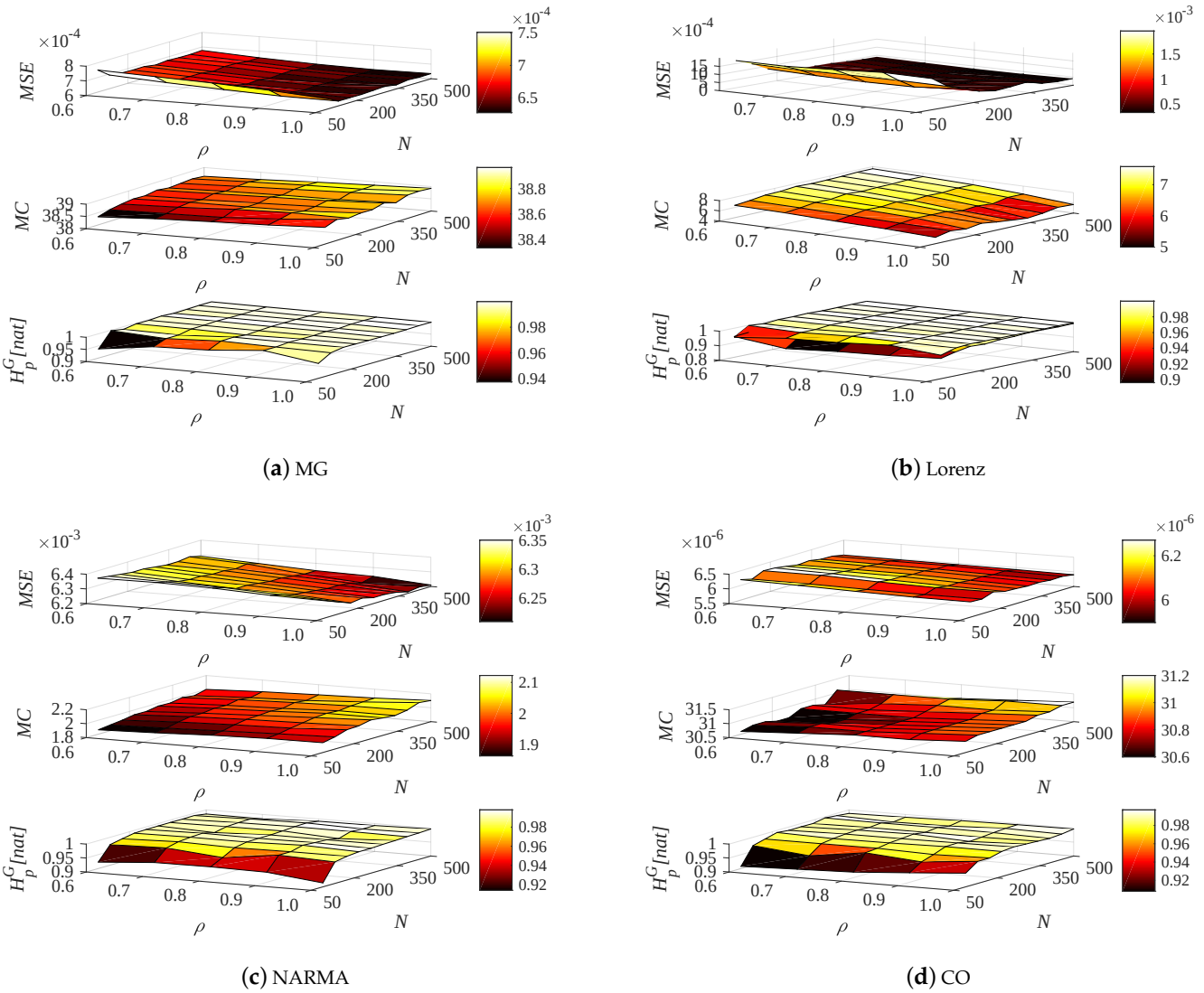
$$MC_k = \frac{Cov^2(u(t-k), y(t))}{Var(u(t))Var(y(t))}, \tag{19}$$

where Cov denotes covariance, Var denotes variance, and $u(t)$ and $y(t)$ denote the input and output of the ESN at moment t, respectively.

Figure 12 evaluates the effectiveness of ESN for the hyperparameters N and ρ in the considered time series prediction tasks. Obviously, there is a very good consistency between the ESN performance (MSE and MC) and the reservoir richness measured by H_p^G via the whole parameter spaces of N and ρ. Concretely, the larger N can yield the lower MSE but the higher MC and H_p^G. Furthermore, Table 2 shows the measure between Pearson correlation H_p^G and ESN performance for different time series tasks, especially considering multi-step ahead prediction mode. From the table, it is obvious that H_p^G has an extremely strong correlation with the MSE and MC over the parameter space of N. Even if the prediction step is up to 5, such a high correlation still remains. It implies that in all settings of N, the reservoir with rich dynamics can be guaranteed, thus yielding good nonlinear projection capability. Given this, the PE theory is feasible for interpreting reservoir projection. On the other hand, from the view of spectral radius, the entropy–performance correlation is significantly weakened. It is due to the fact that the spectral radius just plays a decisive role in the convergence of reservoir internal weight matrix W, ensuring the echo state property, thereby having a slight effect on the richness of the reservoir.

Table 2. Measure on Pearson correlation between H_p^G and ESN performance in the MG, Lorenz, NARMA and CO prediction tasks, where the change steps of N and ρ are 50 and 0.1, respectively.

Step	Parameter Space	Correlation	MG	Lorenz	NARMA	CO
1	$N \in [50, 500]$	H_p^G vs. MSE	−0.96069	−0.95188	−0.85053	−0.69467
		H_p^G vs. MC	0.928394	0.971848	0.791282	0.931725
	$\rho \in [0.1, 1]$	H_p^G vs. MSE	−0.47532	−0.41578	−0.52174	−0.39389
		H_p^G vs. MC	0.588914	0.189191	0.376211	0.689503
3	$N \in [50, 500]$	H_p^G vs. MSE	−0.91901	−0.89865	−0.8507	−0.79467
		H_p^G vs. MC	0.939181	0.921805	0.841936	0.92253
	$\rho \in [0.1, 1]$	H_p^G vs. MSE	−0.56666	−0.348145	−0.56681	−0.41327
		H_p^G vs. MC	0.466588	0.27456	0.59542	0.35036
5	$N \in [50, 500]$	H_p^G vs. MSE	−0.92581	−0.93273	−0.94301	−0.86987
		H_p^G vs. MC	0.927302	0.920522	0.921338	0.91253
	$\rho \in [0.1, 1]$	H_p^G vs. MSE	−0.635644	−0.45148	−0.18493	−0.23585
		H_p^G vs. MC	0.53558	0.554063	0.231826	0.666532



(**a**) MG (**b**) Lorenz

(**c**) NARMA (**d**) CO

Figure 12. ESN performance versus reservoir size N and spectral radius ρ in the MG, Lorenz, NARMA and CO prediction tasks.

Finally, Table 3 shows the MSEs and GSEs for different tasks as the scale increases. In this table, we can find that as the scale increases, the H_p^G decreases while the prediction error of the model increases. This implies that the reservoir has a more powerful nonlinear projection capacity in the single-scale case. On the other hand, the different H_p^G and MSE appear for the considered prediction tasks. These two comments are actually consistent with the single scale and input-driven findings from Figures 9–11.

Table 3. GSE and MSE of ESN at different scales in the MG, Lorenz, NARMA and CO prediction tasks.

s	MG		Lorenz		NARMA		CO	
	MSE	H_p^G	**MSE**	H_p^G	**MSE**	H_p^G	**MSE**	H_p^G
1	6.33×10^{-4}	0.973	3.53×10^{-4}	0.968	6.22×10^{-3}	0.975	2.73×10^{-4}	0.973
2	1.14×10^{-3}	0.945	6.59×10^{-4}	0.93	6.44×10^{-3}	0.945	2.77×10^{-4}	0.944
3	1.20×10^{-3}	0.933	6.52×10^{-4}	0.902	6.68×10^{-3}	0.89	2.96×10^{-4}	0.906
4	1.26×10^{-3}	0.892	7.77×10^{-4}	0.901	6.76×10^{-3}	0.888	3.04×10^{-4}	0.889
5	1.30×10^{-3}	0.872	8.82×10^{-4}	0.871	6.78×10^{-3}	0.863	3.15×10^{-4}	0.835
6	1.36×10^{-3}	0.827	1.06×10^{-3}	0.813	6.79×10^{-3}	0.827	3.26×10^{-4}	0.83

6. Conclusions

In this work, we gain insight into the excellent performance of RC from a PE interpretable framework. In particular, we define two metrics for reservoir analysis (ISE and GSE), which serve as a link between reservoir richness and reservoir projection capability. In addition, multiscale complexity–entropy tools are used to explore the dynamics of reservoir states and neuron-level states at different scales. The simulation results demonstrate the positive correlation between reservoir richness and reservoir projection capacity as well as the single-scale and input-driven properties of the reservoir. In future work, we would like to consider PE to analyze the richness of some deep reservoir structures (e.g., DESN) for the projection interpretability. Moreover, the mutual information derived from entropy theory can be used to capture the correlation of reservoir neurons, assisting the design of better ESNs.

Author Contributions: Conceptualization, X.S. and Y.L.; Software, M.H.; Formal analysis, Y.W. (Yutong Wang); Visualization, Y.W. (Yu Wang); Funding acquisition, Z.L.; Investigation, X.S.; Methodology, X.S. and Y.L.; Validation, X.S. and M.H.; Writing—Original draft, M.H.; Writing—Review & editing, X.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of Hebei Education Department grant number ZD2021088, and by the S&T Major Project of the Science and Technology Ministry of China, Grant 2017YFE0135700.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets for this study are available upon request from the corresponding author.

Acknowledgments: We thank the editorial board and all reviewers for their professional advises to improve this work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Abbreviations and notations in the paper:

Abbreviations	**Description**
ESN	Echo state network
PE	Permutation entropy
ISE	Instantaneous state entropy
GSE	Global state entropy
RC	Reservoir computing
RP	Recurrence plots
ASE	Average state entropy
SCM	Statistical complexity measure
MCPE	Multiscale complexity permutation entropy
MPE	Multiscale permutation entropy
CECP	Complexity–entropy causal plane
MG	Mackey–Glass
NARMA	Nonlinear autoregressive moving average
CO	Crude oil
MSE	Mean square error
MC	Memory capacity
Notation	**Description**
$x(t)$	The state of RC at time t
x_n	The nth neuron of RC
W	Internal weight matrix of ESN
N, n	The number of RC neurons
T, t	Time
X	Internal state matrix of the reservoir

Y	Output matrix of ESN
m	Embedding dimension
τ	Delay time
H_p^t	The state entropy of all neurons in the reservoir at time t
H_p^n	The time-varying state entropy of nth neuron
H_p^I	The value of ISE
H_p^G	The value of GSE
s	The scale factor
ρ	The spectral radius of reservoir
C_{js}	The value of SCM

References

1. Gauthier, D.J.; Bollt, E.; Griffith, A.; Barbosa, W.A. Next generation reservoir computing. *Nat. Commun.* **2021**, *12*, 5564. [CrossRef]
2. Wang, L.; Su, Z.; Qiao, J.; Deng, F. A pseudo-inverse decomposition-based self-organizing modular echo state network for time series prediction. *Appl. Soft Comput.* **2022**, *116*, 108317. [CrossRef]
3. Na, X.; Han, M.; Ren, W.; Zhong, K. Modified BBO-Based Multivariate Time-Series Prediction System With Feature Subset Selection and Model Parameter Optimization. *IEEE Trans. Cybern.* **2022**, *52*, 2163–2173. [CrossRef]
4. Wang, Z.; Zeng, Y.R.; Wang, S.; Wang, L. Optimizing echo state network with backtracking search optimization algorithm for time series forecasting. *Eng. Appl. Artif. Intell.* **2019**, *81*, 117–132. [CrossRef]
5. Xu, M.; Yang, Y.; Han, M.; Qiu, T.; Lin, H. Spatio-Temporal Interpolated Echo State Network for Meteorological Series Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 1621–1634. [CrossRef]
6. Jalalvand, A.; Demuynck, K.; De Neve, W.; Martens, J.P. On the application of reservoir computing networks for noisy image recognition. *Neurocomputing* **2018**, *277*, 237–248. [CrossRef]
7. Tong, Z.; Tanaka, G. Reservoir Computing with Untrained Convolutional Neural Networks for Image Recognition. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 1289–1294.
8. Sun, L.; Jin, B.; Yang, H.; Tong, J.; Liu, C.; Xiong, H. Unsupervised EEG feature extraction based on echo state network. *Inf. Sci.* **2019**, *475*, 1–17. [CrossRef]
9. He, Q.; Pang, Y.; Jiang, G.; Xie, P. A Spatio-Temporal Multiscale Neural Network Approach for Wind Turbine Fault Diagnosis With Imbalanced SCADA Data. *IEEE Trans. Ind. Inform.* **2021**, *17*, 6875–6884. [CrossRef]
10. Cabessa, J.; Hernault, H.; Kim, H.; Lamonato, Y.; Levy, Y.Z. Efficient Text Classification with Echo State Networks. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8.
11. Zhang, Y.; Tiňo, P.; Leonardis, A.; Tang, K. A Survey on Neural Network Interpretability. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *5*, 726–742. [CrossRef]
12. Fan, F.L.; Xiong, J.; Li, M.; Wang, G. On Interpretability of Artificial Neural Networks: A Survey. *IEEE Trans. Radiat. Plasma Med. Sci.* **2021**, *5*, 741–760. [CrossRef]
13. Bianchi, F.M.; Livi, L.; Alippi, C. Investigating Echo-State Networks Dynamics by Means of Recurrence Analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 427–439. [CrossRef]
14. Lee, G.C.; Loo, C.K. On the Post Hoc Explainability of Optimized Self-Organizing Reservoir Network for Action Recognition. *Sensors* **2022**, *22*, 1905. [CrossRef]
15. Bianchi, F.M.; Livi, L.; Alippi, C.; Jenssen, R. Multiplex visibility graphs to investigate recurrent neural network dynamics. *Sci. Rep.* **2017**, *7*, 44037. [CrossRef]
16. Ceni, A.; Ashwin, P.; Livi, L. Interpreting Recurrent Neural Networks Behaviour via Excitable Network Attractors. *Cogn. Comput.* **2020**, *12*, 330–356. [CrossRef]
17. Variengien, A.; Hinaut, X. A journey in ESN and LSTM visualisations on a language task. *arXiv* **2020**, arXiv:2012.01748.
18. Armentia, U.; Barrio, I.; Ser, J.D. Performance and Explainability of Reservoir Computing Models for Industrial Prognosis. In Proceedings of the 16th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2021), Bilbao, Spain, 22–24 September 2022; pp. 24–36.
19. Barredo Arrieta, A.; Gil-Lopez, S.; Laña, I.; Bilbao, M.N.; Del Ser, J. On the post-hoc explainability of deep echo state networks for time series forecasting, image and video classification. *Neural Comput. Appl.* **2022**, *34*, 10257–10277. [CrossRef]
20. Alao, O.; Lu, P.Y.; Soljacic, M. Discovering Dynamical Parameters by Interpreting Echo State Networks. *Present. Neurips Sci. Workshop Dec.* **2021**. Available online: https://openreview.net/forum?id=coaSxusdBLX (accessed on 4 September 2022)
21. Baptista, M.L.; Goebel, K.; Henriques, E.M. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. *Artif. Intell.* **2022**, *306*, 103667. [CrossRef]
22. Han, X.; Zhao, Y. Reservoir computing dissection and visualization based on directed network embedding. *Neurocomputing* **2021**, *445*, 134–148. [CrossRef]
23. Miao, W.; Narayanan, V.; Li, J.S. Interpretable Design of Reservoir Computing Networks Using Realization Theory. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–11. [CrossRef]
24. Tian, Y.; Wang, Z.; Lu, C. Self-adaptive bearing fault diagnosis based on permutation entropy and manifold-based dynamic time warping. *Mech. Syst. Signal Process.* **2019**, *114*, 658–673. [CrossRef]

25. Zunino, L.; Soriano, M.C.; Rosso, O.A. Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. *Phys. Rev. E* **2012**, *86*, 046210. [CrossRef]
26. Ma, F.; Fan, Q.; Ling, G. Complexity-Entropy Causality Plane Analysis of Air Pollution Series. *Fluct. Noise Lett.* **2022**, *21*, 2250011. [CrossRef]
27. Zheng, J.; Pan, H.; Yang, S.; Cheng, J. Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2018**, *99*, 229–243. [CrossRef]
28. à Mougoufan, J.B.B.; Fouda, J.A.E.; Tchuente, M.; Koepf, W. Adaptive ECG beat classification by ordinal pattern based entropies. *Commun. Nonlinear Sci. Numer. Simul.* **2020**, *84*, 105156. [CrossRef]
29. Yin, Y.; Shang, P.; Ahn, A.C.; Peng, C.K. Multiscale joint permutation entropy for complex time series. *Phys. Stat. Mech. Its Appl.* **2019**, *515*, 388–402. [CrossRef]
30. Ozturk, M.C.; Xu, D.; Principe, J.C. Analysis and Design of Echo State Networks. *Neural Comput.* **2007**, *19*, 111–138. [CrossRef]
31. Gallicchio, C.; Micheli, A. Architectural richness in deep reservoir computing. *Neural Comput. Appl.* **2022**, 1–18. [CrossRef]
32. Silva, A.S.A.; Menezes, R.S.C.; Rosso, O.A.; Stosic, B.; Stosic, T. Complexity entropy-analysis of monthly rainfall time series in northeastern Brazil. *Chaos Solitons Fractals* **2021**, *143*, 110623. [CrossRef]
33. Zhang, B.; Shang, P.; Zhou, Q. The identification of fractional order systems by multiscale multivariate analysis. *Chaos Solitons Fractals* **2021**, *144*, 110735. [CrossRef]
34. Pessa, A.A.; Ribeiro, H.V. Ordpy: A Python package for data analysis with permutation entropy and ordinal network methods. *Chaos* **2021**, *31*, 063110. [CrossRef]
35. Herteux, J.; Räth, C. Breaking symmetries of the reservoir equations in echo state networks. *Chaos Interdiscip. J. Nonlinear Sci.* **2020**, *30*, 123142. [CrossRef]
36. Mastroeni, L.; Vellucci, P. Replication in Energy Markets: Use and Misuse of Chaos Tools. *Entropy* **2022**, *24*, 701. [CrossRef]
37. Li, J.; Shang, P.; Zhang, X. Financial time series analysis based on fractional and multiscale permutation entropy. *Commun. Nonlinear Sci. Numer. Simul.* **2019**, *78*, 104880. [CrossRef]
38. Xu, M.; Han, M. Adaptive Elastic Echo State Network for Multivariate Time Series Prediction. *IEEE Trans. Cybern.* **2016**, *46*, 2173–2183. [CrossRef]
39. Yusoff, M.H.; Chrol-Cannon, J.; Jin, Y. Modeling neural plasticity in echo state networks for classification and regression. *Inf. Sci.* **2016**, *364*, 184–196. [CrossRef]
40. Li, D.; Liu, F.; Qiao, J.; Li, R. Structure optimization for echo state network based on contribution. *Tsinghua Sci. Technol.* **2018**, *24*, 97–105. [CrossRef]