

RESEARCH

Open Access



# Design of a reservoir for cloud-enabled echo state network with high clustering coefficient

Abbas Akrami<sup>1</sup>, Habib Rostami<sup>2\*</sup> and Mohammad R. Khosravi<sup>2,3</sup>

## Abstract

Reservoir computing (RC) is considered as a suitable alternative for descending gradient methods in recursive neural networks (RNNs) training. The echo state network (ESN) is a platform for RC and nonlinear system simulation in the cloud environment with many external users. In the past researches, the highest eigenvalue of reservoir connection weight (spectral radius) was used to predict reservoir dynamics. Some researchers have illustrated; the characteristics of scale-free and small-world can improve the approximation capability in echo state networks; however, recent studies have shown importance of the infrastructures such as clusters and the stability criteria of these reservoirs as altered. In this research, we suggest a high clustered ESN called HCESN that its internal neurons are interconnected in form of clusters. Each of the clusters contains one backbone and a number of local nodes. We implemented a classical clustering algorithm, called K-means, and three optimization algorithms including genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) to improve the clustering efficiency of the new reservoir and compared them with each other. For investigating the spectral radius and predictive power of the resulting reservoirs, we also applied them to the laser time series and the Mackey-Glass dynamical system. It is demonstrated that new clustered reservoirs have some specifications of biologic neural systems and complex networks like average short path length, high clustering coefficient, and power-law distribution. The empirical results illustrated that the ESN based on PSO could strikingly enhance echo state property (ESP) and obtains less chaotic time series prediction error compared with other works and the original version of ESN. Therefore, it can approximate nonlinear dynamical systems and predict the chaotic time series.

**Keywords:** Reservoir computing, Echo state networks, Complex networks, Clustering, Time series prediction, Scale-free analysis

## 1 Introduction

Unlike feed-forward neural networks, it is costly and challenging to train recurrent neural networks with traditional methods such as gradient descent in the presence of feedback loops. The echo state network (ESN) is considered to be a suitable alternative for gradient descent algorithms [1] in for cloud-based services [2]. Given that in ESN, both the input and internal connections weight

matrix are constant, but the network output matrix is trained by the linear regression method; this efficient training method causes ESN has a highly dynamic behavior with little learning complexity. However, the random structure of the reservoir may reduce its accuracy of the estimation. Nevertheless, many effective schemes of reservoir production have proposed. Some works use the analytical method to construct the reservoirs but use these methods to solve the complex problems or discrete functions is difficult. To further improve the reservoir's performance, some other works are looking for evolutionary techniques [3–5]. Furthermore, echo state property

\* Correspondence: [habib@pgu.ac.ir](mailto:habib@pgu.ac.ir)

<sup>2</sup>Computer Engineering Department, Faculty of Engineering, Persian Gulf University, Bushehr, Iran

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

(ESP) [6] is the astonishing capabilities in ESN. As mentioned by Jaeger [7], only provided the current reservoir state is exclusively specified by the long-time history of inputs after running; the ESN has echo state property (ESP). Such a way that, for ESN, the maximum eigenvalue of reservoir connection matrix, spectral radius, must be no larger than 1 in such a way that the ESP is maintained [8]. It should be noted that the bigger the spectral radius, the slower the network response to the input pulses, and the network memory capacity is enhanced. Hence, the ESN can have a better approximation ability and a more efficient computing power. In some sense, the spectral radius has an extraordinary effect on the approximation abilities of the ESN. To guarantee ESP without the reservoir weight scale, a reservoir production method has been proposed using the eigenvalue decomposition, and the convergence of the algorithm is also theoretically guaranteed. However, eigenvalues are still generated randomly [9].

In recent years, small-world phenomenon and scale-free property in many complex systems of real-world, like immune systems, biological networks, transport networks, internet backbone, citation networks, and many other networks have been discovered [10–17]. The small-world feature refers to the short characteristic path length as well as the high clustering coefficient in the network. It is noteworthy that the complex artificial neural networks such as associative memory systems [18] using the small-world phenomenon, and the scale-free property [19, 20] have better efficiency in time and memory capacity than the randomly connected networks with similar connections. Also, the chaotic time series prediction problem by the small-world trait and the scale-free property can be solved more efficiently [21–24]. It is noteworthy that the forecast precision of a chaotic time series using ESN substantially increases with a ratio of 2400 compared with former methods [25].

Deng and Zhang [26] suggested a complex ESN model with a gradual growth state reservoir called HCESN, which included a lot of internal nodes with sparse connections. Their experimental results showed that the echo state property could be improved by permitting a large-scale spectrum of the acceptable spectral radius.

Jarvis et al. [27] investigated the impact of clusters, hierarchies, and interconnections between clusters on the spectral radius prediction ability. Their experimental results showed that hierarchy and almost small cluster size increments the amplitude of the spectral radius in the reservoir. Also, they showed that the size of the entire reservoir and the connection between the clusters effect on the allowed range of spectral radius. As well as, for clustering reservoir neurons in [28] classical clustering algorithms have been used.

In this article, we suggest four new (ESNs) and compare them with each other and previous works towards cloud computing. Then, the best network is selected as the desired reservoir. We showed that four suggested reservoirs have scale-free property, small-world trait, and distributed structure. Neurons in the first network are clustered using the K-means clustering technique, and the resulting reservoir is called HCESN-KM. The second network is clustered by the genetic algorithm (GA), and the created reservoir is named HCESN-GA. In the third network, clustering is done using the algorithm of differential evolution (DE), and the resulting reservoir is called HCESN-DE. And in the fourth network, neurons are clustered by particle swarm optimization (PSO) and are named HCESN-PSO. In each cluster, the meaningful nodes are considered as backbone neurons and other nodes as local neurons. Hence, connections between neurons have small-world topology and follow a power-law distribution so that the resulting models can reflect the learning mechanism of most biological systems. We also applied the new reservoirs to challenging difficulties like Mackey-Glass (MG) and laser time series prediction problems and evaluated them with each other and with previous works. The empirical results show that the suggested methods, particularly HCESN-PSO, outperform the previous ones in terms of the capability of approximating nonlinear dynamic systems and prediction accuracy of chaotic time series.

Other sections include the following. In Section 2, the classical clustering and three evolutionary clustering algorithms are briefly described. In Section 3, the proposed new state reservoirs using the small-world and the scale-free topologies are explained, and in Section 4, the complexities of HCESNs are analyzed. Dynamic approximation ability and enhanced echo state property are investigated and compared with previous works in Section 5. In the long run, the last section is devoted to conclusions.

## 2 Related work

### 2.1 Classical clustering

Clustering is an unsupervised learning issue because it classifies unlabeled data into classes or clusters, in such a way that the data within the same clusters have the most similarity and between different clusters have the most difference. Hence, it is needed to define a measurement criterion for these similarities and established a benchmark for allocating data to particular cluster centers. One of these criteria is the Euclidean distance of two data  $x$  and  $y$  or  $\text{dist}(x, y)$ . The smaller the interval between  $x$  and  $y$ , the higher the similarity between them and vice versa [29]. This method attempt to assign the data in the dataset  $D$  to  $k$ -clusters  $C_1, \dots, C_k$  so that  $C_i \subset$

$D$ , and  $C_i \cap C_j = 0$  for  $1 \leq i, j \leq k$ . The clustering quality of  $C_i$  is measured as follows:

$$\text{MSE} = \sum_{i=1}^k \sum_{p \in C_i} \text{dist}(p, c_i)^2 \quad (1)$$

Here, MSE is the mean square error between all data in  $C_i$  and the centroid of  $c_i$ .

Although the K-means clustering algorithm is a simple and well-known method, it may get stuck at local optimum solutions, depending on the selection of the primary cluster centers. To overcome such a challenging problem, many evolutionary computation techniques have been suggested. In the next section, the proposed clustering algorithms are described. We employ these algorithms for clustering the internal neurons of proposed HCESNs in Section 3 [29].

## 2.2 Genetic algorithm-based clustering technique

Genetic algorithm (GA) is the initial conjecture-based search and optimization algorithm that is inspired by biological evolution. The algorithm searches for multiple paths simultaneously and hence reduces the probability of getting stuck in the local optimum solutions.

A cluster similarity metric, or a fitness function, can be used to search for optimal cluster centers in the feature space. To use the genetic algorithm to solve optimization problems, unknown variables are encoded in string structures. Each string is called a chromosome. Chromosomes are considered to encode many fixed cluster centers [29]. A set of strings (chromosomes) is called a population. At first, a population is created randomly that denote different objects. Each of them is related to an objective or fitness function that is optimized based on the principle of genetics and natural evolution. Inspired by biological factors like crossover and mutation, a new generation of the chromosome is produced. The mentioned process is to continue until conditions are met [30].

## 2.3 Differential evolution (DE)

This algorithm is a multipurpose optimization method that can find almost optimal solutions to real and mathematical problems. Individuals in differential evolution are indicated by  $d$ -dimensional vectors  $v_i$ ,  $i \in [1, \dots, np]$ , where  $d$  is the number of nonlinear optimization operators, and  $np$  is the population size. This method is proposed to overcome the local search problem in the genetic algorithm. The main difference between GA and DE is the order of *crossover* and *mutation* operators, also in how the *selection* operator works. According to [31], the classic evolutionary process of DE is as follows. For more details, see [32, 33].

### 2.3.1 Generation of initial population

Using uniform distribution, the value of each individual is selected within the range  $[v_j^{\min}, v_j^{\max}]$ .

$$v_i(j) = v_j^{\min} + \text{unifrnd}(0, 1) \left( v_j^{\max} - v_j^{\min} \right) \quad (2)$$

$\text{unifrnd}(0, 1)$  function is used to generate a random number with  $[0, 1]$  using uniform distribution.

### 2.3.2 Mutation

The mutation process delivers vector  $y_i$  as an outcome by applying a strategy like  $y_i = v_{i1} + F \cdot (v_{i2} - v_{i3})$

Randomly select three individual  $v_{i1}$ ,  $v_{i2}$ ,  $v_{i3}$ , from the current generation where  $i_1 \neq i_2 \neq i_3$  are an integer,  $F > 0$  is the mutation factor and controls the difference of the mutation  $di = v_{i2} - v_{i3}$ .

### 2.3.3 Crossover

In the crossover, to get the trial vector  $u_i(j)$ , an element is taken either from donator vector  $y_i(j)$  or target vector  $v_i(j)$  based on the following expression

$$u_i(j) = \begin{cases} y_i(j), & \text{if } u_j(0, 1) < CR \\ v_i(j), & \text{otherwise} \end{cases} \quad (3)$$

where  $u_j(0, 1)$  denotes the uniform random distribution among  $(0, 1)$ , and CR is the rate of crossover.

### 2.3.4 Selection

If get vector, the individual of the next generation is replaced.

$$v'_i = \begin{cases} u_i, & \text{if } f(u_i) < f(v_i) \\ v_i, & \text{otherwise} \end{cases} \quad (4)$$

where  $v'_i$  means the offspring of  $v_i$  for the later generation [31].

## 2.4 Particle swarm optimization (PSO)

This algorithm is an evolutionary calculation method and a population-based stochastic search process introduced by Kenney and Eberhart (1995) [34]. The PSO models the social behavior of a group of birds or fish and is a subset of swarm intelligence. Each particle represents a *position* and a *velocity* in  $Q$ -dimensional search space  $G$ , and it adjusts its position to the best particle position ever (*pbest*) and the best position in the particle population (*gbest*). Initially, the positions and velocities of all individuals are randomly determined. In each phase, first, the particle *velocity* and then its *position* are updated. Hence, each of the particles has a memory keeping their best

position. The *velocity* and *position* of the particles are adjusted as follows:

$$v_{iq}(t) = \phi v_{iq}^{(t-1)} + \omega_1 r_{1q}^{(t-1)} (pbest_{iq}^{(t-1)} - p_{iq}^{(t-1)}) + \omega_2 r_{2q}^{(t-1)} (gbest_{iq}^{(t-1)} - p_{iq}^{(t-1)}) p_i(t) = p_i^{(t-1)} + v_i^{(t-1)} \quad (5)$$

where  $\phi$  is the inertia weight,  $\omega_1$  and  $\omega_2$  are the acceleration constants, and both  $r_1$  and  $r_2$  are uniform random distribution in the range of  $[0, 1]$ .

In this section, we consider three optimization algorithms, such as GA, DE, and PSO, as well as the K-means clustering algorithm. We apply these algorithms to several standard datasets from the University of California Irvine (UCI) database like Iris, Wine, CMC, Bupa, Vowel, Cancer, and Thyroid and analyze their results. Here, the Euclidean distance is used to the sum of the mean square error. The algorithms mentioned above have separately implemented 30 times for each UCI dataset by the Matlab 2016 software. The result is illustrated in Table 1.

The comparison results show that the PSO and K-means algorithms have the lowest and the highest cost functions in six datasets, respectively. On the other hand, the DE algorithm in the Wine dataset has the least cost function among the six datasets, and the GA has less cost function than the DE algorithm in the Bupa and Vowel datasets. The results in Table 1 show that PSO clustering methods have lots of potentials. As well as, in the general case, Fig. 1 shows the cost function for the K-means, GA, DE, and PSO algorithms with validity indices CS [35] and DB [36].

### 3 Proposed method

#### 3.1 HCEAN's reservoir architecture

The reservoir architecture is illustrated in Fig. 2, which includes input units ( $I$ ), internal units ( $N$ ), and output units ( $Q$ ).

The network activations for input units at the time step  $t$  are  $U(t) = [u_1(t), u_2(t), \dots, u_i(t)]^T$  through an  $N \times I$  input weight matrix  $W^{in}$ . The reservoir nodes are sparsely connected, and the output of each internal node is known as a state and denoted by  $X(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$

through an  $N \times N$  internal reservoir weight matrix  $W^{res}$ . The output layer is collected all the inputs  $u_i(t)$  in the layer one, and all the states  $x_i(t)$  in the new state network via a  $Q \times (N \times I)$  output connection matrix  $W^{out}$ , and in the long run, generate an output vector  $y(t) = [y_1(t), y_2(t), \dots, y_q(t)]^T$  of the HCESNs. At the same time, the output vector  $y(t)$  has feedback to whole the internal nodes via an  $n \times q$  connection weight matrix  $W^{back}$ . The activation function of  $\tanh()$  is set for the last two layers. The weight matrices of input and feedback  $W^{in}$  and  $W^{back}$  are determined using uniform random distribution, as well as, the output connection weight matrix  $W^{out}$  is tuned with supervised learning. Even so, the reservoir connection matrix  $W^{res}$  is generated corresponding to our proposed evolutionary rules instead of the merely random methods used in [7, 25]. The readout  $y(t)$  of HCESNs is implemented as follows:

$$x(t) = \tanh(W^{res}x(t-1) + W^{in}u(t) + W^{fb}y(t-1) + v(t-1))$$

$$y(t-1) = \tanh\left(W^{out}\begin{bmatrix} x(t-1) \\ u(t-1) \end{bmatrix}\right) \quad (6)$$

where  $v(t)$  is a threshold noise. Production of new state reservoirs includes

1. First, create a  $H \times H$  grid space and initialize it.
2. Generate backbone neurons in the grid space by the uniform distribution.
3. Produce synaptic relationships using the clustering methods for recently added local neurons. Here, a classical clustering algorithm like K-means and three metaheuristic optimization algorithms including genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for clustering the internal neurons are used. The mean local neurons in each cluster are called the backbone unit. The number of backbone neurons is approximately 1% of local neurons [26]. The backbone neurons coordinate on the grid space are randomly created. And also, the resulting domains were separated. The distance among the backbone neurons should not be less than a specified basis. Adding the new local neurons to the

**Table 1** Comparison of four proposed algorithms on the UCI dataset

Algorithms	Criteria	UCI dataset						
		Iris	Wine	CMC	Bupa	Vowel	Cancer	Thyroid
K-means	MSE	97.27	18090	321.3	5543.99	152501.2	3056.4	1980.3
GA	MSE	96.67	16298	312.5	5532.5	149121	3041.9	1891.5
DE	MSE	96.67	16296	312.5	5638.22	160171.2	3041.9	1882.1
PSO	MSE	96.67	16298	312.5	5532.21	148971	3041.7	1867.5

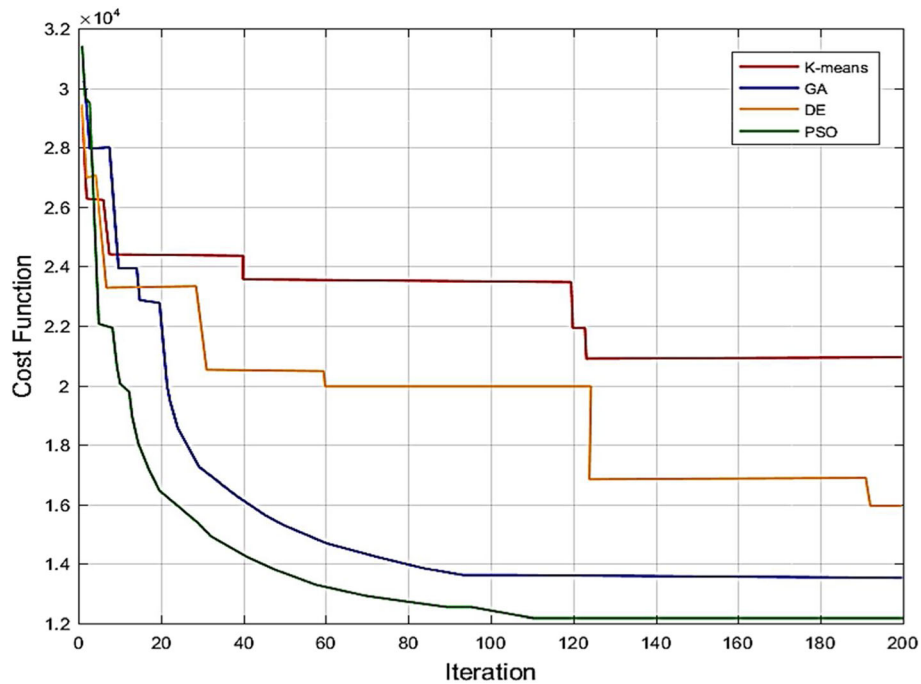


Fig. 1 Comparison graph of four proposed algorithms using DB and CS index

reservoir space creates a fully connected grid of the backbone neurons  $N_{backbone}$ . In this process, we randomly selected the coordinates  $(X, Y)$  of one of the backbone neurons and placed the coordinates  $x$  and  $y$  of the local neurons associated with the backbone of the grid space to distribute the unbalanced power law of outdegree using the algorithms mentioned above. This spatial distribution of neurons is very similar to the function of the human brain system [37, 38].

- Applying preferential attachment rules [11]. The rule causes the neurons that recently added are attached to the neurons that have the most synaptic connectivity. If the coordinates of the

new local neuron are considered as the center of the circle where the environment is the location of the backbone neurons, then the radius of that circle is regarded as the Euclidean distance. As a result, the backbone neurons have the most distance with new local neurons in the candidate neighbors. Our preferential attachment rules include the following steps:

Assuming that the  $N_{link}$  is the number of links for new local neurons. Also,  $n_{candidate}$  and  $n_{current}$  are the number of neurons in the candidate region and the current region of a new local neuron, respectively. ( $n_{candidate} \leq n_{current}$ ).

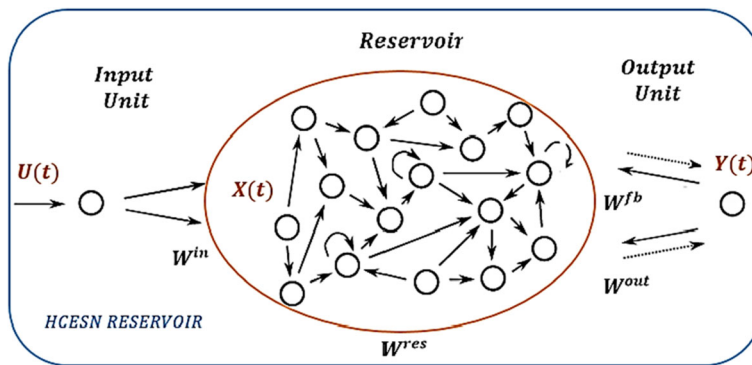


Fig. 2 HCESN's architecture

1. If  $N_{\text{link}} \geq n_{\text{current}}$ , so, the link among the new local neurons to all neurons in the current region is complete.
2. If  $N_{\text{candidate}} \leq N_{\text{link}} < N_{\text{current}}$ , then the link among the new local neurons to all candidate neighborhood is generated with the probability of  $O_i / \sum_{j \in C_{\text{neighbor}}} O_j$  [39] where  $O_i$  and  $O_j$  are the outdegrees of neurons  $i$  and  $j$ , and  $C_{\text{neighbor}}$  is the candidate neighborhood for new neurons.

If  $n_{\text{connection}} < n_{\text{candidate}}$ , the process of connecting candidate neurons to local neurons is the same as before.

Generally speaking, candidate neurons can help improve network clustering coefficient and preferential attachment rules to both small-world and scale-free features.

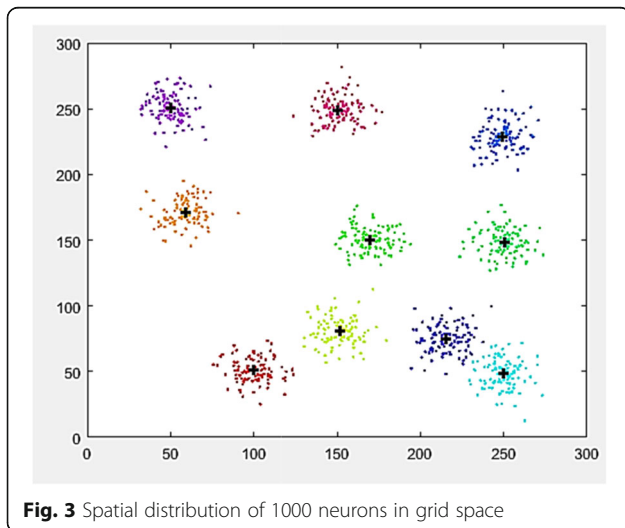
## 4 Experimental results and discussion

### 4.1 Computational complexity

The HCESN topology with the reservoir capacity of  $Q = 300 \times 300$  is generated according to the following parameters: the number of interior nodes  $N_{\text{interior}} = 1000$ , the links of local nodes  $N_{\text{link}} = 5$ , and the number of backbone nodes  $N_{\text{backbone}} = 10$ . The spectral radius of this new reservoir is 0.979%, and its sparse connection is 2.105%.

### 4.2 Distributed and hierarchical architecture

As illustrated in Fig. 3, the new reservoir contains 1000 internal neurons in a  $300 \times 300$  grid and 10 clear clusters or domains. Each of the color dots denotes an interior neuron created with the normal growth rules and clustering methods. In this paper, DB and CS clustering indices are used for clear clustering [40]. In each cluster, the backbone neurons are surrounded by a set of local neurons. The clusters are considered as macro-neurons at the top of the reservoir network hierarchy, and the



**Fig. 3** Spatial distribution of 1000 neurons in grid space

number of interconnections within clusters is usually much greater than those that are between clusters. Therefore, dynamic behaviors are relatively independent of each cluster. The input connection matrix  $W^{\text{in}}$  between input and interior neurons is determined using a uniform random distribution, and those input connections that are not linked to the interior neurons of the network are converted to zero. Therefore, the connection weight distribution of the input in the reservoir is a spatial distribution. Hence, our network architecture has a distributed and hierarchical spatial structure.

### 4.3 Small-world property

In graph theory, the clustering coefficient  $C$  and the minimum characteristic path length  $L$  are used to describe the phenomenon of the small-world [10, 41]. The reservoir network that is growing naturally consists of  $10^3$  internal neurons. Accordingly, it has a  $10^3 \times 10^3$  reservoir connection matrix  $W^{\text{res}}$ , by the sparse connection of 0.979%. Therefore, it is a complex and extensive network. In general, the minimum characteristic path length denotes the gauge of a graph or complex system. As well as, it is referred to as the average interval overall pairs between two interior nodes.

The average short characteristic path length is specified as  $L = \binom{N}{2}^{-1} \sum_{(i \neq j)} l_{ij}$ . In which  $N$  is the number of neurons,  $l_{ij}$  is the minimum interval between neurons  $i$  and  $j$ , and  $\binom{N}{2}$  demonstrated all possible pairs of neurons [42]. The average short characteristic path length for HCESN-KM, HCESN-GA, HCESN-DE, and HCESN-PSO reservoirs was computed as  $L_{\text{HCESN-KM}} = 3.4011$ ,  $L_{\text{HCESN-GA}} = 3.3801$ ,  $L_{\text{HCESN-DE}} = 3.1795$ , and  $L_{\text{HCESN-PSO}} = 3.1401$ .

The clustering coefficient is determined as the mean part of pairs of neighbor neurons of an interior node that are adjacent to each other. The clustering coefficient is computed by  $C = 1/N \sum_{i=1}^N C_i$  where  $C_i$  is the clustering coefficient for node  $i$  specified as  $C_i = 2E_i / K_i(K_i - 1)$ . Here,  $E_i$  is the real edges of the neighbor node  $i$ ,  $K_i$  is the total neighbors connected to the neuron  $i$ , and  $K_i(K_i - 1)/2$  is the maximum number of possible connectivities between the neighbor neurons. Hence,  $C_i$  denotes a ratio of real neighbor neuron connectivity to the maximum possible connectivities [42]. The clustering coefficient for HCESN-KM, HCESN-GA, HCESN-DE, and HCESN-PSO reservoirs was computed as  $C_{\text{HCESN-KM}} = 0.4727$ ,  $C_{\text{HCESN-GA}} = 0.4811$ ,  $C_{\text{HCESN-DE}} = 0.4827$ , and  $C_{\text{HCESN-PSO}} = 0.4916$ .

Deng and Zhang [26] represented a reservoir named SHESN. For this network, they computed the average path length  $L = 3.7692$  and the clustering coefficient  $C = 0.2303$ . As well as, they calculated a random system with a similar size. For this network, they computed the small

characteristic path length and the clustering coefficient, respectively,  $L_R = 32668$  and  $C_R = 0.0112$ .

Our experimental results denote that the short characteristic path length  $L$  related to four new reservoirs is approximately as near as  $L_{\text{random}}$ , and the clustering coefficients  $C$  are immensely more extensive than  $C_{\text{random}}$ . Hence, our new HCESNs are the complex network with the small-world phenomenon.

Besides, Table 2 illustrates the short characteristic path length and coefficient of clustering for each of the ten clusters in four networks. According to the table, it is observed that each cluster is also a sub-network of the small-world.

#### 4.4 Scale-free property

Recently, it has been observed that the specific features of internet topologies can be described using the power law in the form of  $y = x^\gamma$  [43], where  $\gamma$ , as an index or degree of the power law, is used to describing some of the properties of the global network topologies. So that, reservoirs with power-law distribution are called scale-free [26]. Deng and Zhang [26] observed the gradient of the resulting linear plot between  $(x, y)$  on a log-log scale as the power law of  $\gamma$ . They also used Pearson's correlation coefficient ( $p$ ) to ensure that the power law existed. The closer the correlation coefficient is to 1, the more the data follow the power-law distribution [39].

We investigate two types of power-law distributions as follows: the number of nodes against outdegree and out-degree of nodes against rank. It should be noted that the rank of a node is defined as the number of connections of one node to another and can be calculated using the *corrcoef* in MATLAB. The correlation coefficients for HCESN-KM, HCESN-GA, HCESN-DE, and HCESN-PSO reservoirs with the  $p$  value of 0, were calculated as 0.987, 0.982, 0.986, and 0.984 respectively. Also, the correlation coefficient relationship between outdegree and number of nodes for HCESN-KM, HCESN-GA, HCESN-DE, and HCESN-PSO reservoirs was calculated as 0.968, 0.971, 0.978, and 0.989, respectively. As well as, we calculated the correlation coefficient for each of the 10 clusters. As shown in Table 2, for each cluster as a low-level sub-networks, power-law properties exist. Hence, the proposed HCESNs have some biological features, such as scale-free distribution [38].

#### 4.5 Supervised learning

As mentioned in Section 3, to maintain the echo state property, the connection matrix  $W^{\text{res}}$  of the new reservoirs must be attentively selected. And the input and the feedback connections ( $W^{\text{in}}$  and  $W^{\text{fb}}$ ) could be arbitrarily determined using the uniform distribution. The output connection matrix  $W^{\text{out}}$  using supervised learning must be adjusted. In such a way that the RNN with the echo

**Table 2** The characteristics of the scale-free and small-world for ten clusters

Clusters	01	02	03	04	05	06	07	08	09	10
Reservoir type	HCESN-KM									
Cluster size	121	115	99	92	103	86	88	108	85	103
Average path length	2.0370	2.0713	1.9376	1.9741	1.9987	1.9586	1.9321	2.0119	1.9836	1.9989
Clustering coefficient	0.3955	0.4001	0.4828	0.4767	0.4097	0.4495	0.4698	0.4711	0.4247	0.4092
Correlation coefficient	0.9851	0.9897	0.9848	0.9860	0.9881	0.9898	0.9824	0.9875	0.9802	0.9825
Reservoir type	HCESN-GA									
Cluster size	94	102	86	110	93	95	127	108	103	82
Average path length	1.9801	1.8813	1.9616	2.0751	1.9668	1.9696	2.0833	2.0041	1.8874	1.8931
Clustering coefficient	0.4604	0.4101	0.4186	0.4369	0.4867	0.4491	0.4554	0.4798	0.4366	0.4287
Correlation coefficient	0.9889	0.9834	0.9811	0.9799	0.9871	0.9810	0.9900	0.9839	0.9834	0.9812
Reservoir type	HCESN-DE									
Cluster size	98	105	104	101	78	116	93	117	85	103
Average path length	1.9789	2.0854	2.0156	1.8809	1.8857	2.0015	1.9411	2.0048	1.9607	1.8787
Clustering coefficient	0.3454	0.4605	0.4251	0.4249	0.4357	0.4491	0.4854	0.3981	0.4257	0.4602
Correlation coefficient	0.9801	0.9800	0.9802	0.9851	0.9891	0.9901	0.9889	0.9813	0.9864	0.9860
Reservoir typ	HCESN-PSO									
Cluster size	83	107	99	103	121	84	99	108	87	109
Average path length	1.9121	2.0103	1.9526	1.8214	1.9678	1.9423	1.9512	1.9943	1.8837	1.9949
Clustering coefficient	0.4221	0.4787	0.4874	0.4609	0.4112	0.4425	0.4824	0.4842	0.4214	0.4782
Correlation coefficient	0.9827	0.9881	0.9953	0.9910	0.9871	0.9890	0.9959	0.9893	0.9872	0.9901

state property can approximate the following sample dataset with the length of  $n_r$ .

$$\{u(n_1), y_d(n_1)\} \cdot \{u(n_2), y_d(n_2)\} \dots \{u(n_r), y_d(n_r)\} \quad (7)$$

After throwing away the initial transition  $n_0$ , we must find  $W^{\text{out}}$  to reduce the mean square error (MSE).

$$\text{MSE} = 1 / (n_r - n_0) \sum_{t=n_0+n_1}^{n_r} \left( d(t) - W^{\text{out}} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \right)^2 \quad (8)$$

where  $u(t)$  and  $y_d(t)$  is input and desired output vectors at the time( $t$ ), respectively. It should be noted that  $d(t) = \tanh^{-1} y_d(t)$ ,  $x(t) = [x_1(t), x_2(t) \dots x_n(t)]^T$  and ( $n$ ) denotes the echo state parameters. An inverse matrix is used to predict this linear regression model. Hence, the matrix  $W^{\text{out}}$  is obtained as follows:

$$W^{\text{out}} = (M_G^{-1} D)^T \quad (9)$$

where  $T$  is the transpose. The  $(N + 1) \times (n_r - n_0)$  dimensional matrix is given as follows:

$$M_G = \begin{bmatrix} x_1(n_0 + n_1) \dots & x_n(n_0 + n_1) & u(n_0 + n_1) \\ \vdots & \ddots & \vdots \\ x_1(n_r) \dots & x_n(n_r) & u(n_r) \end{bmatrix} \quad (10)$$

$$D = [d(n_0 + n_1) \dots d(n_r)]^T \quad (11)$$

It should be noted that the generalized inverse matrix computation ( $M_G$ ) has been performed with the *pinv* function in MATLAB [26].

## 5 Test criteria

### 5.1 Assessment of HCESNs capabilities

#### 5.1.1 Mackey-Glass dynamic system

Mackey-Glass (MG) dynamic system, with a long time-delay  $\delta$ , is an appropriate test-bed for the prediction of nonlinear chaotic systems [26, 44–46]. This system has been used by a challenging problem to verify the efficiency of the new HCESNs. The MG differential equation is given by

$$\frac{ds}{dt} = \frac{0.2 \times s(t-\delta)}{1 + s(t-\delta)^{10}} - 0.1s(t) \quad (12)$$

where  $s$  and  $\delta$  denote the state and the time-delay, respectively. A chaotic time series occurs in the Mackey-Glass when  $\delta \geq 17$ . The reason for the turbulence of this system is that the slightest change in the initial

conditions has the most significant impact on the system output.

Also, for solving differential equations with constant delay, the function of the *dde23* in MATLAB, which contains a set of training and test data, is used. In particular, instead of the predefined value of the *dde23* function, we determine absolute precision with  $(1e-16)$ . To compare, we perform our experiments, according to datasets supported by Jaeger and Haas [25] and as well as Deng and Zhang [26] (Table 3).

#### 5.1.2 Laser time series

The laser time series is broadly employed to testing a variety of prediction methods of the real-world chaotic time series [25, 44, 47–49]. We used datasets 18 and 19 as used by Deng and Zhang [26].

## 5.2 Formulations of the problem

At this stage, the accuracy of HCESNs is evaluated in 100 independent runs, either at an identified observation point or for whole specified points. In particular, for the MG system, 84 instances from the learning dataset by length  $n_k$  were selected as follows:  $\{u(n_r + 1); y_d(n_r + 1)\}$ ,  $u(n_r + 2); y_d(n_r + 2)\}$  to  $\{u(n_r + 84); y_d(n_r + 84)\}$ .

$$\text{NRMSE}_{84} = \sqrt{\frac{\sum_{i=1}^{100} (y_d^i(n_r + 84) - y^i(n_r + 84))^2}{\sum_{i=1}^{100} (y_d^i(n_r + 84))^2}} \quad (13)$$

where  $y^i(t)$  denotes the network output in the  $i$ th test to predict the laser time series problem. Also, we considered the observation point  $t = n_r + 84$  and performed 100 separate experiments. To compute the normalized root mean square error (*NRMSE*). This time, we use all 200 data points in the experimental dataset, namely  $n_t + 1$  to  $n_t + 200$ , to calculate *NRMSE* for 100 unique implementations.

$$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^{100} \sum_{m=n_r+1}^{n_r+200} (y_d^i(m) - y^i(m))^2}{\sum_{i=1}^{100} \sum_{m=n_r+1}^{n_r+200} (y_d^i(m))^2}} \quad (14)$$

**Table 3** Mackey-Glass training and testing datasets

Sample datasets	( $\delta$ ) Delay	Conversion
1	17	$\tanh(s - 1)$
2	30	$(0.3 \tanh(s - 1)) + 0.2$
3-17	17-31	$(0.3 \tanh(s - 1)) + 0.2$



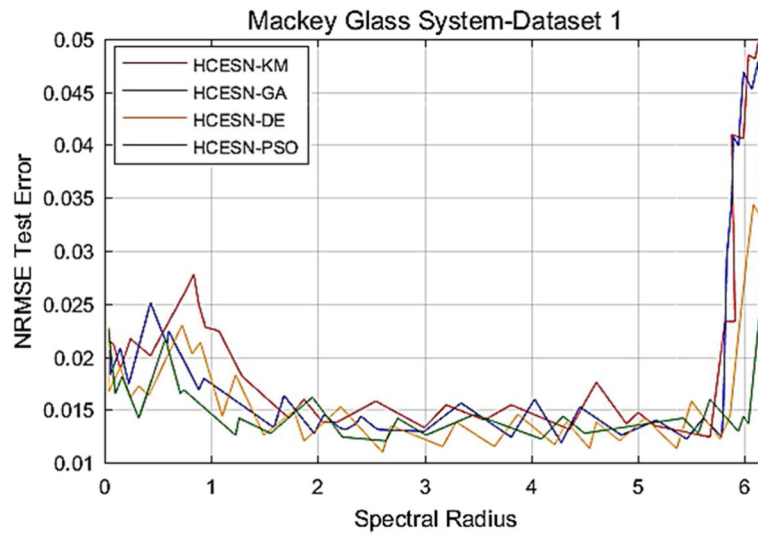


Fig. 4 NRMSE84 test error on dataset 1

### 5.3 Enhanced echo state property

According to the definition given by Jaeger in [25], if the maximum eigenvalue of the reservoir connection-weight matrix  $|\lambda_{\max}(W^{res})|$  or spectral radius is no more than one, then ESN has echo state. It is fully compatible with the experimental studies of Deng and Zhang [26], that is, when the network has a spectral radius of more than one, the ESN is not able to work correctly. But they succeeded in enhancing the echo state property by increasing the spectral radius till the boundary of 6.0 in SHEEN.

Here, we performed 4 experiments using the datasets 1, 2, 18, and 19 on the Mackey-Glass system and laser time series prediction to check the echo state properties of

the HCESNs. We calculated the NRMSE84 exam errors by increasing the spectral radius by the step size of 0.1. Our empirical results are illustrated in Figs. 6, 7, and 6, respectively. The following parameters are considered for testing HCESNs in datasets 1 and 2.

Reservoir capacity  $H \times H = 200 \times 200$ , number of internal nodes  $n_{in} = 500$ , number of new local node connections  $n_{local} = 5$ , number of backbone nodes  $n_{backbone} = 5$ , input connection-weight  $W^{in}$ , and feedback connection  $W^{fb}$  were tuned by uniform distribution from  $-1$  to  $1$ . The output connection-weight  $W^{out}$  was obtained by supervised learning. The empirical results in the Mackey-Glass system for HCESNs are illustrated in Figs. 4 and 5, respectively.

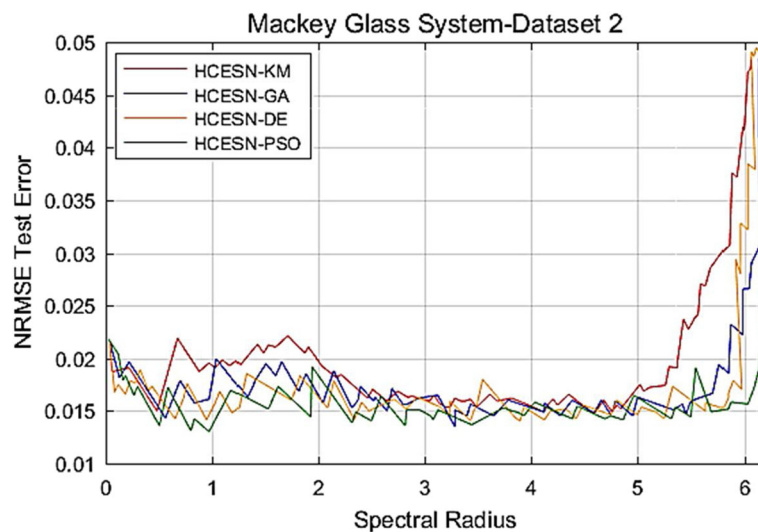
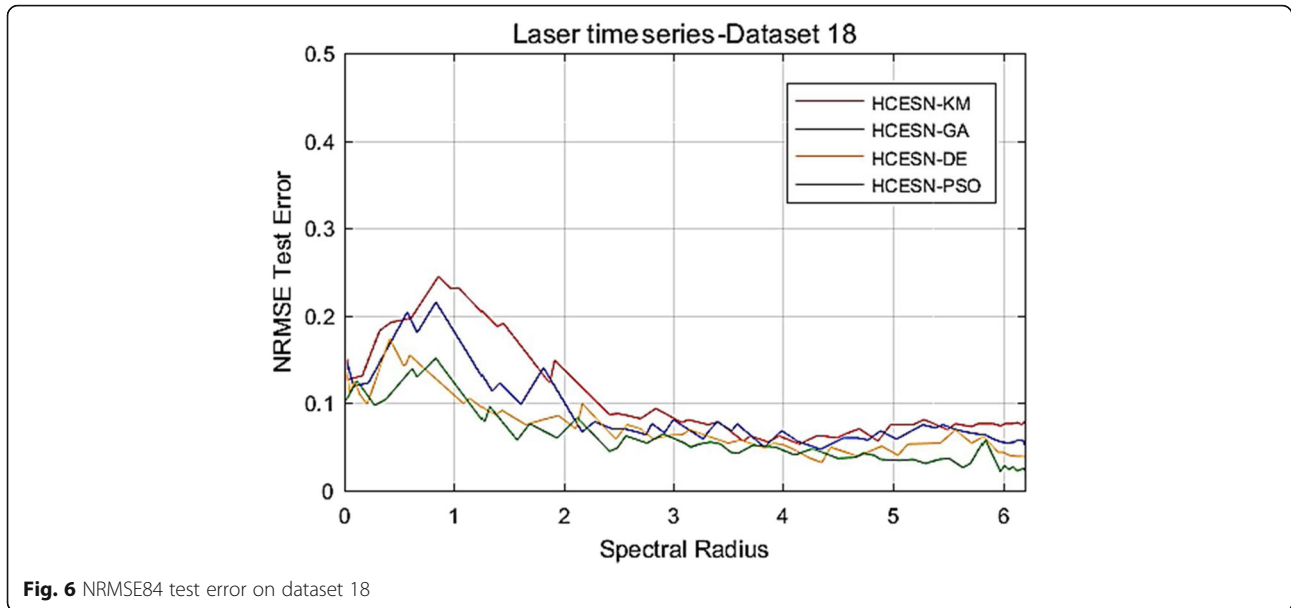


Fig. 5 NRMSE84 test error on dataset 2



The parameters used in datasets 18 and 19 are the same as those used in datasets 1 and 2, except that the number of new local neuron connections  $n_{\text{local}} = 1$ , input connection weight  $W^{\text{in}}$ , and feedback connection  $W^{\text{fb}}$  were tuned by uniform distribution from  $-1$  to  $1$ . Our empirical results in the Mackey-Glass system for HCESNs are illustrated in Figs. 6 and 7, respectively.

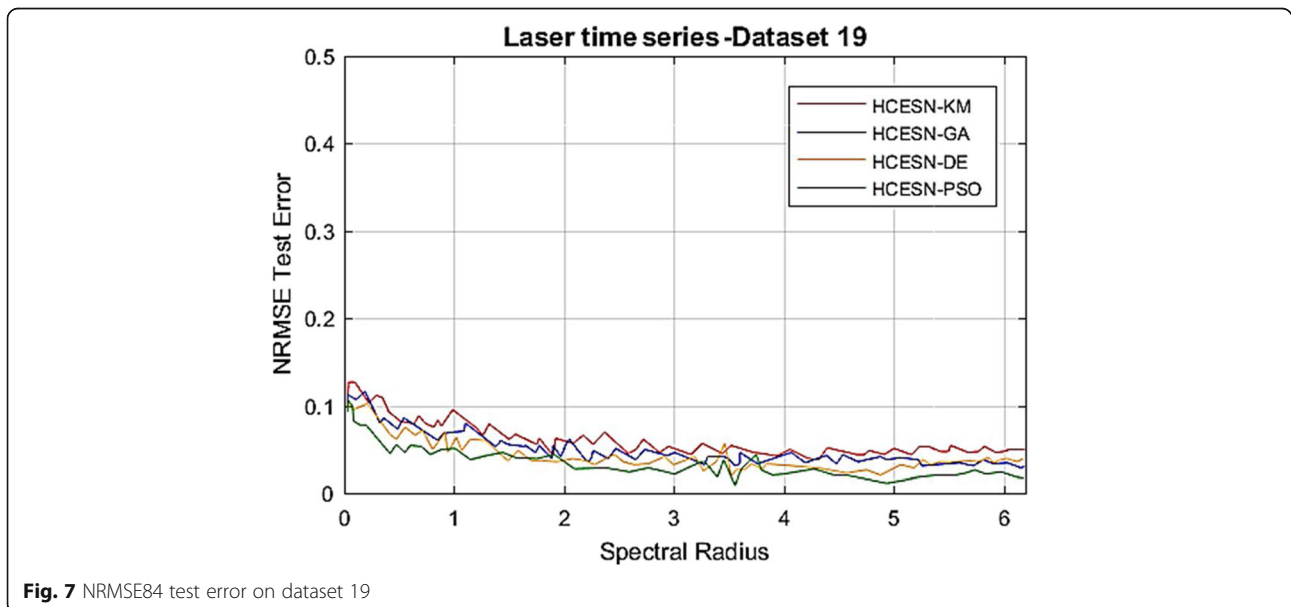
The results demonstrate that the HCESN based on PSO clustering is capable of having a more extensive range of the spectral radius than the other three networks discussed in this paper. Even at a spectral radius

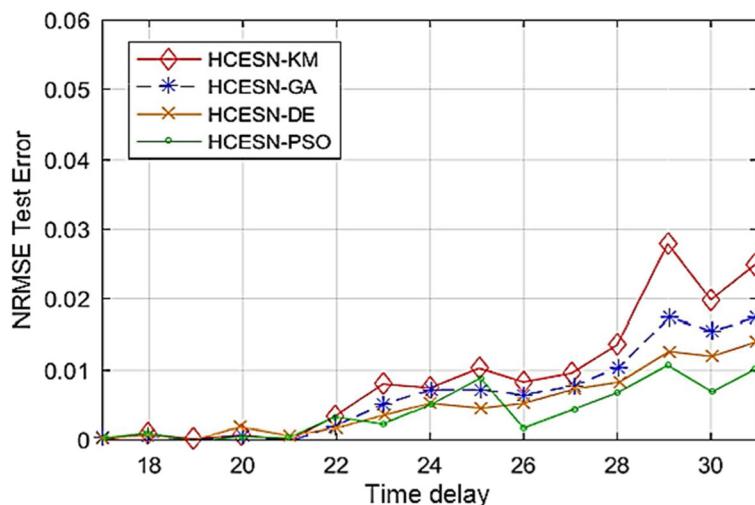
more significant than 1, this reservoir can significantly enhance the echo state property.

#### 5.4 HCESN's capability to nonlinear approximation

##### 5.4.1 Mackey-Glass system

As mentioned above, for an MG system, by increasing time-delay  $\delta$ , the system becomes extremely nonlinear. In particular, a chaotic time series occurs when  $\delta \geq 17$ . Therefore, the approximation of the Mackey-Glass system using the echo state network or any other model with increasing  $\delta$  is almost impossible, and it is undoubtedly a significant challenge. Deng and Zhang [26]





**Fig. 8** The NRMSE test error against time-delay  $\delta$  for HCESNs in MG

applied the SHESN dynamic nonlinear approximation capability first to model a dynamic MG system with a high delay. To compare SHESN with ESN, they used datasets 3–17, which included variable time delays of 17 to 31, respectively. In Table 5, the standard deviation for our HCESNs and other ESNs are listed. By increasing  $\delta$  to 26 and beyond, the nonlinear approximation of ESN dramatically increases, which means that the nonlinearity dynamics of the system are much more critical, and it is very complicated to deal with this problem. For instance, when  $\delta = 29$ , the SHESN and new HCESNs have a suitable performance. However, HCESN-PSO has higher accuracy than SHESN (Fig. 8).

#### 5.4.2 Laser time series prediction

In Figs. 6 and 7, the prediction capability of HCESNs was evaluated based on datasets 18 and 19. Then, the prediction accuracy of HCESNs with SHESN in the spectral radius of 4.0 was compared, as mentioned by Deng and Zhang [26]. The results were also compared with the prediction accuracy of the entirely random ESN reservoir with a spectral radius of 0.9 in [26]. In Table 4, the list of NRMSE test errors in 100 independent runs is shown.

The standard deviation for ESN, SHESN, and HCESNs is also shown in Table 5 for delays of 17 to 31.

As shown for ESN in [7, 25], if the spectral radius  $|\lambda_{\max}(W^{\text{res}})|$  is greater than 1, the echo state property (ESP) does not continue. But Deng and his colleagues [26] have shown that in the natural growing SHESN network, by permitting a wide scope of the acceptable spectral radius, the ESP increases.

According to [26], the largest eigenvalue in ( $i = 1$ ) for both ESN and SHESN is considered 1.5. The slope of the curve corresponding to the ESN reservoir is very slow, and the size of each 100 eigenvalues is about 1.5. At the same time, SHESN reservoir eigenvalues are heavily decreased. And only 1.2% of all eigenvalues are larger than 1. However, the distribution of all eigenvalues follows the power law, in which the correlation coefficient with  $p = 0$  is 0.989. We also achieved this fantastic phenomenon at HCESNs. So, we calculated the average of the total eigenvalues in 100 independent runs. Experimental results confirm that magnitudes of maximum eigenvalues are less than 1 and follow the power law. The correlation coefficients for HCESN-KM, HCESN-GA, HCESN-DE, and HCESN-PSO were 0.9671, 0.986, 0.988, and 0.992, respectively, as shown in Fig. 9.

## 6 Conclusions

In this study, an enhanced ESN was proposed, which includes a classic clustering algorithm and three

**Table 4** NRMSE test error for prediction accuracy of laser time series

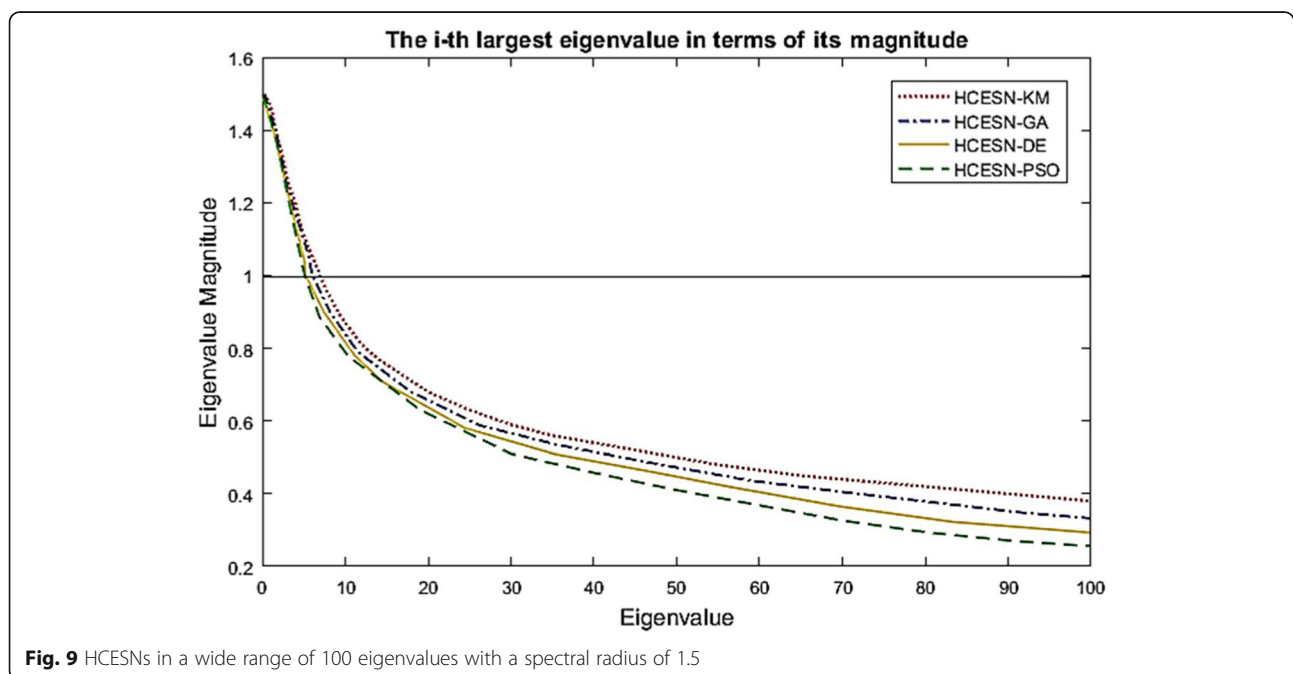
Network type	Dataset 18	Dataset 19
Original ESN	0.1422	0.0804
Natural growth SHESN	0.1148	0.0558
HCESN-KM	0.1140	0.0510
HCESN-GA	0.1138	0.0497
HCESN-DE	0.1132	0.0501
HCESN-PSO	0.1107	0.0442

**Table 5** Standard deviation for Mackey-Glass system

Time delay $\delta(\times 10^{-4})$	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Original ESN	0.032	0.159	0.053	0.181	0.068	0.276	0.398	0.117	0.794	0.233	0.271	0.624	2.551	0.968	2.755
Natural Growth SHESN	0.037	0.149	0.031	0.268	0.048	0.384	0.364	0.171	1.769	0.171	0.337	0.847	0.689	0.889	2.191
HCESN-KM	0.038	0.136	0.021	0.189	0.035	0.375	0.353	0.169	1.365	0.138	0.315	0.738	0.588	0.811	2.121
HCESN-GA	0.037	0.149	0.061	0.235	0.058	0.380	0.369	0.172	1.475	0.175	0.338	0.891	0.690	0.891	2.195
HCESN-DE	0.039	0.148	0.057	0.221	0.042	0.375	0.385	0.138	1.487	0.179	0.327	0.898	0.694	0.910	2.194
HCESN-PSO	0.036	0.141	0.045	0.215	0.069	0.371	0.391	0.168	1.451	0.169	0.331	0.824	0.701	0.881	2.178

evolutionary optimization algorithms called as HCESN-KM, HCESN-GA, HCESN-DE, and HCESN-PSO. Results showed that PSO-based clustering is faster and has a lower cost function than the other proposed clustering. Therefore, HCESN-PSO is recommended for reservoir design. As well as, results in Section 4.3 denote that HCESN-PSO has the shortest characteristic path length and the highest clustering coefficient, which includes the scale-free property and small-world phenomenon. We investigated two kinds of power-law distribution, such as the number of nodes vs. outdegree and outdegree of nodes vs. rank. We presented several natural incremental growth rules that include such as (a) average path length, (b) high clustering coefficient, (c) scale-free property, (d) distributed and hierarchical architectures. We reviewed all the behaviors of HCESNs and applied them to the Mackey-Glass system and the laser time series prediction. The empirical results confirm that, compared with the utterly random ESN by Jaeger [7], as well as

the SHESN proposed by Deng and Zhang, our new HCESNs networks, specially HCESN-PSO, which include thousands of neurons or even more, can significantly enhance the echo state property (ESP) and approximate the highly complex nonlinear dynamic systems. Such an efficient system is likely to represent some biological neural properties, such as the small-world phenomenon and scale-free distribution. In order to applications in other areas of research and applied developments, we suggest applications of the proposed method in environmental and energy studies which uses soft computing techniques [50–52]. This study also tried to make the HCESNs architecture more robust against noise on specific (hub) neurons. Undoubtedly, optimization methods with mathematical proofs and accurate statistical analysis of improved echo state property of ESNs are some of the most exciting and important issues that will be investigated in the future.



### Abbreviations

RC: Reservoir computing; RNN: Recursive neural networks; ESN: Echo state network; GA: Genetic algorithm; DE: Differential evolution; PSO: Particle swarm optimization; ESP: Echo state property; UCI: University of California Irvine; MG: Mackey-Glass

### Acknowledgements

The authors thank the honorable reviewers and editors for their valuable comments, handling, and suggestions on this manuscript.

### Authors' contributions

AA participated in the mathematical design of the proposed method and its computer implementation. HR and MK coordinated the industrial application and raw data preparation and helped out for the study. AA, HR, and MK have completed the first draft of this paper. All authors have read and approved the final manuscript.

### Authors' information

Not applicable.

### Funding

Not applicable.

### Availability of data and materials

All the data and computer programs are available.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Information and Communication Technology Engineering Department, Faculty of Engineering, Payame Noor International University of Assaluyeh, Bushehr, Iran. <sup>2</sup>Computer Engineering Department, Faculty of Engineering, Persian Gulf University, Bushehr, Iran. <sup>3</sup>Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran.

Received: 11 December 2019 Accepted: 17 February 2020

Published online: 13 March 2020

### References

1. W.S. McCulloch, W. Pitts, A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull Math Biol.* **5**, 115–133 (1943) <https://doi.org/10.1007/BF02478259>
2. J. Kim, H.J.T. Manaligod, J. Lee, S. Jo, *Cloud Networking Computing* (2019)
3. S. Otte, M.V. Butz, D. Koryakin, F. Becker, M. Liwicki, A. Zell, Optimizing recurrent reservoirs with neuro-evolution. *Neurocomputing.* **192**, 128–138 (2016)
4. Chouikhi, N., Ammar, B., Rokbani, N., Alimi, A. M., & Abraham, A. (2015). A hybrid approach based on particle swarm optimization for echo state network initialization. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on* (pp. 2896-2901). IEEE.
5. J. Chen, D. Liu, F. Hao, H. Wang, Community detection in the dynamic signed network: an intimacy evolutionary clustering algorithm. *J Ambient Intelligence Human Comp.*, 1–10 (2019)
6. I.B. Yildiz, H. Jaeger, S.J. Kiebel, Re-visiting the echo state property. *Neural Net.* **35**, 1–9 (2012)
7. Jaeger, H. (2001). The “echo state” approach to analyzing and training recurrent neural networks with an erratum Note. Bonn, Germany: German National Research Center for Information Technology GMD Technical The report, 148(34), 13.
8. M. Buehner, P. Young, A tighter bound for the echo state property. *IEEE Trans Neural Netw* **17**(3), 820–824 (2006)
9. J. Qiao, F. Li, H. Han, W. Li, Growing Echo-State Network With Multiple Subreservoirs. *IEEE Trans. Neural Netw. Learning Syst.* **28**(2), 391–404 (2017)
10. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks nature, 393(6684), 440.
11. Barabási, A. L., & Albert, R. (1999). The emergence of scaling in random networks.science, 286(5439), 509-512.
12. M.R. Khosravi, S. Samadi, Reliable Data Aggregation in Internet of VISA R Vehicles Using Chained Dual-Phase Adaptive Interpolation and Data Embedding. *IEEE Internet of Things Journal.* (2019)
13. A.L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization. *Nat Rev Gen.* **5**(2), 101 (2004)
14. Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review* (Vol. 29, No. 4, pp. 251-262). ACM.
15. K. Klemm, V.M. Eguiluz, Highly clustered scale-free networks. *Physical Review E* **65**(3), 036123 (2002)
16. S.H. Strogatz, Exploring complex networks. *Nature.* **410**(6825), 268 (2001)
17. J. Travers, S. Milgram, The small world problem. *Psychol Today.* **1**(1), 61–67 (1967)
18. Yang, J., He, L., & Kong, B. (2016). Efficient Method for Designing Associative Memory with Contextual small-world Architecture. In *2016 9th International Symposium on Computational Intelligence and Design (ISCID)* (Vol. 2, pp. 152-156). IEEE.
19. D.H. Kim, J. Park, B. Kahng, Enhanced storage capacity with errors in scale-free Hopfield neural networks: An analytical study. *PLoS one.* **12**(10), e0184683 (2017)
20. Umamaheshwari, S., & Swaminathan, J. N. (2018, January). Man-In-Middle Attack/for a scale-free Topology. In *2018 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-4). IEEE.
21. F. Han, M. Wiercigroch, J.A. Fang, Z. Wang, Excitement and synchronization of small-world neuronal networks with short-term synaptic plasticity. *Int J Neural Syst* **21**(05), 415–425 (2011)
22. C. Li, Q. Zheng, Synchronization of the small-world neuronal network with unreliable synapses. *Phys Biol.* **7**(3), 036010 (2010)
23. Y. Tang, F. Qian, H. Gao, J. Kurths, Synchronization in complex networks and its application—a survey of recent advances and challenges. *Ann Rev Control.* **38**(2), 184–198 (2014)
24. Alderisio, F., & di Bernardo, M. (2018). Controlling the collective behavior of networks of heterogeneous Kuramoto oscillators with phase lags. In *2018 European Control Conference (ECC)* (pp. 2248-2253). IEEE.
25. Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication.science, 304(5667), 78-80.
26. Z. Deng, Y. Zhang, Collective behavior of a small-world recurrent neural system with scale-free Distribution. *IEEE Transac Neural Networks.* **18**(5), 1364–1375 (2007)
27. Sarah J, Stefan R, Ulrich E (2010) Extending stability through hierarchical clusters in Echo State Networks, *Frontiers in Neuroinformatics*, 2010-volume 4
28. E. Najibi, H. Rostami, SCE SN, SPESN, SWESN: Three recurrent neural echo state networks with Clustered reservoirs for prediction of nonlinear and chaotic time-series. *Applied Intelligence.* **43**(2), 460–472 (2015)
29. U. Maulik, S. Bandyopadhyay, Genetic algorithm-based clustering technique. *Pattern recognition.* **33**(9), 1455–1465 (2000)
30. Sörensen, K., & Glover, F. W. (2013). Metaheuristics. *Encyclopedia of operations research and management science*, 960-970.
31. R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Global Optimization* **11**(4), 341–359 (1997)
32. odder, T., Bhattacharya, D., & Chakraborty, S., Adaptive Differential Evolution with Intersect Mutation and Repaired Crossover Rate. *Int J Comp Intelligence IoT.* **2**(1) (2019)
33. S. Das, P.N. Suganthan, Differential evolution: A survey of the state-of-the-art. *IEEE Trans Evolutionary Computation.* **15**(1), 4–31 (2010)
34. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization (PSO). In *Proc. IEEE International Conference on Neural Networks, Perth, Australia* (pp. 1942-1948).
35. T. Caliński, J. Harabasz, A dendrite method for cluster analysis. *Commun Stat Theory Methods* **3**(1), 1–27 (1974)
36. D.L. Davies, D.W. Bouldin, A cluster separation measure. *IEEE Transac Pattern Anal Machine Intelligence.* **2**, 224–227 (1979)
37. H. Lee, D. Golkowski, D. Jordan, S. Berger, R. Ilg, J. Lee, G. Golmizaie, Relationship of critical dynamics, functional connectivity, and states of consciousness in large-scale human brain networks. *NeuroImage.* **188**, 228–238 (2019)
38. Eguiluz, V. M., Chialvo, D. R., Cecchi, G., Baliki, M., & Apkarian, A. V. (2004) scale-free brain functional networks. *Neuroimage*, 22, 2330.
39. A. Medina, I. Matta, J. Byers, On the origin of power-laws in Internet topologies. *ACM SIGCOMM computer communication review.* **30**(2), 18–28 (2000)

40. S. Das, A. Abraham, A. Konar, Automatic clustering using an improved differential evolution algorithm. *IEEE Transac Syst Man Cybernetics-Part A: Systems and Humans*. **38**(1), 218–237 (2008)
41. Y. Kawai, J. Park, M. Asada, A small-world topology enhances the echo state property and signal propagation in reservoir computing. *Neural Networks*. (2019)
42. Sohn, I. (2017). small-world and scale-free network models for IoT systems. *Mobile Information Systems*, 2017.
43. Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review* (Vol. 29, No. 4, pp. 251–262).
44. H.G. Han, L. Zhang, Y. Hou, J.F. Qiao, Nonlinear model predictive control based on a self-organizing recurrent neural network. *IEEE transactions on neural networks and learning systems* **27**(2), 402–415 (2016)
45. Ni, T., Wang, L., Jiang, Q., Zhao, J., & Zhao, Z. (2018). LSHADE with semi-parameter adaptation for chaotic time-series prediction. In *Advanced Computational Intelligence (ICACI), 2018 Tenth International Conference on* (pp. 741–745). IEEE.
46. M.C. Mackey, L. Glass, Oscillation and chaos in physiological control systems. *Science*. **197**(4300), 287–289 (1977)
47. Chandra, R. (2018, July). Multi-Task Modular Backpropagation For Dynamic time-series Prediction. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–7). IEEE.
48. Weigend, A. S. (2018). *time-series prediction: forecasting the future and understanding the past*. Routledge.
49. L. Aguayo, G.A. Barreto, Novelty Detection in time-series Using Self-Organizing Neural Networks: A Comprehensive Evaluation. *Neural Processing Letters*. **47**(2), 717–744 (2018)
50. B. Safarianejadian, Using Adaptive Neuro Fuzzy Inference System (ANFIS) for Prediction of Soil Fertility for Wheat Cultivation. *Biol Forum*. **9**(1), 37–44 (2017)
51. M.J. Mokarram, Robust and effective parallel process to coordinate multi-area economic dispatch (MAED) problems in the presence of uncertainty, IET Generation. *Trans Distribution* **13**(18), 4197 (2019)
52. M.J. Mokarram, Hybrid Optimization Algorithm to Solve the Nonconvex Multiarea Economic Dispatch Problem. *IEEE Syst J*. **13**(3) (2019)

## Publisher's Note

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

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---