Echo State Wavelet-sigmoid Networks and Their Application to Nonlinear System Identification

Xiaochuan Sun, *Member, IAENG,* Yingqi Li, Jiayu Liu and Minghui Zhang

***Abstract*—Wavelet theory has become popular in modeling echo state network. One of the most promising directions is usage of the wavelets as membership activation functions of its reservoir. However, only Symlets wavelet seems to be suitable for hybrid wavelet-sigmoid activation functions. To enhance a systematic study of the field, we concentrate on developing more wavelets towards the typical ESN representation, and ask: What are the more outstanding wavelets of reservoir structure for obtaining competitive models and what is the memory capacity of such reservoirs for obtaining competitive models? In the paper, three echo state wavelet-sigmoid networks (ESWNs) are proposed, considering Shannon wavelet, frequency B-spline wavelet and impulse response wavelet, respectively. The corresponding wavelet functions are instead of sigmoid one in part to construct wavelet-sigmoid reservoirs. On three widely used system nonlinear approximation tasks of different origin and characteristics, as well as by conducting a theoretical analysis we show that the proposed ESWNs are superior to the popular echo state network with build-in Symlets wavelet.**

***Index Terms*—echo state network, wavelet function, short-term memory capacity, system identification**

I. Introduction

RECURRENT Neural Networks (RNNs) are a widely known class of neural networks used for learning in sequential domains. In theory, RNNs can approach any nonlinear system with arbitrary precision [1][2], and exhibit dynamic temporal pattern, since they introduce recurrent connections between neurons. However, RNN training algorithm [3] based on direct optimization of the network weights is relatively tedious so that it is easy to cause slow convergence, excessive calculation and suboptimal solutions [4].

To settle all the aforementioned issues,a groundbreaking network structure for RNNs was proposed independently by Jaeger, called echo state networks (ESNs) [5]. ESN is a recurrent neural network, whose the core is a large and fixed reservoir. The reservoir is composed of massive randomly and sparsely connected neurons, and passively stimulated by the input signal. The desired readout is the only trainable part, which can be generated by a linear readout layer attached to the reservoir based on a simple linear regression. Typically, the important property of an ESN is the echo state property (ESP): the reservoir state is an echo of the entire input history. ESN can operate steadily if its spectral radius is set to a proper value in the preparation phase, which is completely different from traditional RNNs. ESNs have been successfully applied in various domains, e.g., time-series prediction [6], pattern recognition [7], robot control [8], reinforcement learning [9] and noise modeling [10].

ESNs have attracted the widespread attentions of research communities, and many extensions of the classical ESN have been explored, such as echo state Gaussian process [6], minimum complexity ESN [11], Balanced ESNs [12], tree ESNs [13], deterministic ESN with loop reservoir [14], copula ESN [15]. In these improved ESN paradigms, sigmoid function (S-function) was the most widely used as the activation function of reservoir neurons because of its stability. However, despite its popularity, S-function is not an optimal choice for nonlinear approximation [18].

Recently, wavelet transform [16][17] has become a widespread tool for modeling neural networks. Typical computation advantages that wavelet functions have include the following [19][20]: wavelets have local support and provide compact local representation of signals in both time and frequency domain; wavelets are capable of scaling (width) and shifting (centers), that allow to process signals with local features. In [21], Jain et al., proposed an adaptive wavelet neural network (WNN) for low-order dominant harmonic estimation. The model possessed high estimation accuracy and robustness against the presence of interharmonics, noise, and fundamental frequency deviation. In [22], Bhaskar et al., developed a two-stage forecasting method for multiple wind power prediction, in which adaptive WNNs are used to predict each decomposed signal. It was demonstrated that this method had better approximation and faster training speed than feed-forward neural network. In [23], Wang et al., proposed a wavelet-neuro-fuzzy system for chaotic time series identification. The multidimensional wavelet, instead of conventional radial basis functions, was conducive to the improvement of approximation property and learning rate. Especially in [24], Wang et al., firstly introduced wavelet neurons as an extension of the typical ESN with sigmoid neurons, where wavelet neurons replaced sigmoid ones in part. Furthermore, Cui et al., injected the hybrid wavelet-sigmoid neurons to minimum complexity ESN [25]. These two models enhanced prediction performance and memory capacity. However, in the above wavelet-injected methods, only a few classical wavelet functions is considered towards ESNs, such as Symlets, Morlet, and Gaussian wavelets [26]. Moreover, Symlets wavelet function was regarded as the optimal activation one of reservoir neurons. To broaden a study of the field, we concentrate on exploring more wavelet families to build activation functions in the typical ESN architecture.

Unlike the above popular wavelets, the Shannon wavelet,

Manuscript received March 28, 2017; revised May 16, 2017. This work was supported in part by supported by the National Natural Science Foundation of China under Grant 61503120.

XC. Sun is with the School of Information Engineering, North China University of Science and Technology, Tangshan, 063210 China (corresponding author, phone: +86 18630501637; e-mail: sunxiaochuan@ncst.edu.cn).

YQ. Li is with with the School of Information Engineering, North China University of Science and Technology, Tangshan, 063210 China (e-mail: liyingqi@ncst.edu.cn).

JY. Liu is with with the School of Information Engineering, North China University of Science and Technology, Tangshan, 063210 China.

MH. Zhang is with with the School of Information Engineering, North China University of Science and Technology, Tangshan, 063210 China.

frequency B-spline wavelet and impulse response wavelet have more outstanding advantages in multiresolution representation of signals. The Shannon wavelet not only has orthogonality but also approximate any functions in the quadratic integrable space, leading to good nonlinear mapping ability [27]. The B-spline wavelet manifests the competitive superiorities of explicit expressions in time and frequency domains relative to the capability of nonlinear approximation, and symmetry or antisymmetry conducive to function reconstruction [28]. The impulse response wavelet effectively represents the system behavior while guaranteeing admissibility and providing sufficient smoothness and rate of decay in both time and frequency domains [29]. Theoretically, the excellent characteristics of these wavelets can provide a more favorable treatment for reservoir activation functions.

In the paper, we introduce three wavelet families of SW, FBW and IRW towards ESN, respectively, providing the promising alternatives to SymW. Embedded in the reservoir neurons, we propose three echo state wavelet-sigmoid networks (ESWNs). In these newly built structures, wavelet neurons are partly injected into reservoirs, and have different activation functions that are derived from the same wavelet family. The hybrid activation mechanism, as well as the outstanding local feature representation of alternative wavelet functions, providing better theoretical guarantees that ESWNs will perform well on given nonlinear approximation tasks. Through the experiments with widely used time series benchmarks, we observe that compared with a SymW-based ESN, these three new structures are able to provide efficient solutions to the predictive modeling task.

The contents of this paper is organized in six sections. Section 2 presents our echo state wavelet-sigmoid networks. Experimental results are shown in Section 3. We analyze both theoretically and empirically the short-term memory capacity (MC) of our models in Section 4. Finally, our work is discussed and concluded in Sections 5 and 6, respectively.

II. Echo state wavelet-sigmoid network

This paper focuses on developing new wavelets towards ESN to achieve good approximation quality, termed as echo state wavelet-sigmoid networks. The details of ESWNs are briefly discussed in the following subsections.

A. Selection of wavelet functions

Definition 1: In function space $L^2(\Re)$ (or Hilbert space), the mother wavelet function $h(x)$ must satisfy the following admissibility condition.

$$C_\psi = \int_{\Re} \frac{\left|\widehat{\psi}(\omega)\right|^2}{|\omega|} d_\omega < \infty \tag{1}$$

where $\widehat{\psi}(\omega)$ is the Fourier transform of $h(x)$. The corresponding wavelet family is obtained through the dilation and translation, given by

$$\psi_{a,b}(x) = |a|^{-\frac{1}{2}} \psi(\frac{x-b}{a}) \quad a, b \in \Re, a \neq 0 \tag{2}$$

where a is the dilation factor, b is the translation factor. Eq.(2) is the continuous wavelet function dependent on the parameter (a, b), generated from the mother wavelet $\psi(t)$.

Unlike the common wavelets in ESN modeling, such as Symlets, Morlet and Gaussian wavelets, we consider the alternative wavelets in the paper as follows.

(1) The Shannon wavelet [27] is constructed from the Shannon multiresolution approximation, which approximates functions by the restriction to low frequency intervals. The real SW can be expressed as:

$$\psi^{Shannon}(t) = 2sinc(2t-1) - sinc(t) \tag{3}$$

This wavelet belongs to the C^∞ class of differentiability, but it decreases slowly at infinity and has no bounded support, since band-limited signals cannot be time-limited.

(2) The frequency B-spline wavelet [28] is a series of mother wavelets, each of which produces a different wavelet family, defined as follows

$$\psi^{Bspline}(t) = C_{m,f_b} sinc^m(\frac{f_b t}{m}) e^{j2\pi f_c t} \tag{4}$$

where C_{m,f_b} is a normalization factor of the wavelet energy, $m \in N$ is the integer order parameter, $f_b \in \Re^+$ is the bandwidth parameter, and $f_c \in \Re^+$ is the wavelet's central frequency.

(3) The impulse response wavelet, proposed by Alkafafi et al. in [29], is derived from the bearing impulse response, aimed at increasing the amplitude of the generated wavelet coefficients related to the fault impulses, and improving the fault detection process, given by

$$\psi^{Impulse}(t) = Ae^{-\frac{\beta}{\sqrt{1-\beta^2}}\omega_c t} \sin(\omega_c t) \tag{5}$$

where A is an arbitrary scaling factor, β is the damping factor that adjusts the resolution of the wavelet, and ω is the wavelet centre frequency.

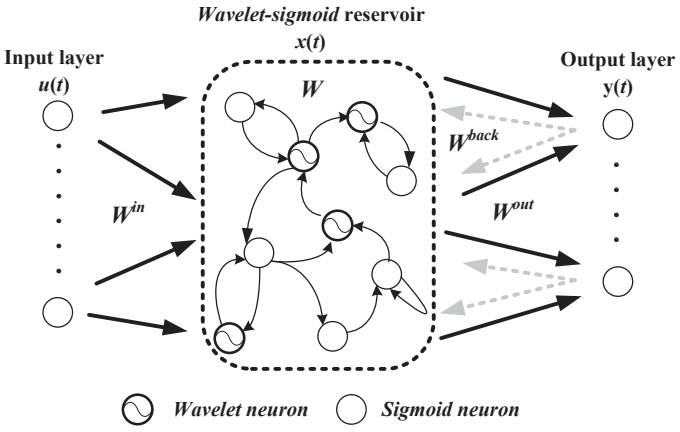


Fig. 1. Typical architecture of an ESWN.

B. Network architecture of ESWN

ESWN is essentially an ESN architecture with a hybrid wavelet-sigmoid reservoir. The wavelets in the Section II-A serve as its reservoir activation functions, and they are first introduced to build the wavelet-sigmoid reservoir. Fig. 1 shows the typical architecture of a multiple-input multiple-output ESWN, comprising of three key parts: input layer, hybrid wavelet-sigmoid reservoir and output layer. $u(t)$ is the observed signal injected to the network at time t, W^{in} is the input weight matrix, i.e., the matrix of connections

between the input units and the internal units, $x(t)$ is the reservoir state at time t, W is the reservoir weight matrix, i.e., the matrix of connections between the internal units, $y(t)$ is the network readout at time t, W^{out} is the output weight matrix, i.e., the matrix of connections from internal units to output units, W^{back} is the feedback weight matrix, i.e., the matrix of connections from output units to internal units. Supervised ESWN training is conducted by updating the reservoir state and network output as follows:

$$x(t+1) = f(W^{in}u(t+1) + Wx(t) + W^{back}y(t)) \quad (6)$$

$$y(t+1) = W^{out}(x(t+1)) \quad (7)$$

where f denotes the activation function of reservoir neurons, given by

$$f(\cdot) = \begin{cases} \tanh(\cdot) & if\ sigmoid \\ \psi_{a,\,b}(\cdot) & if\ wavelet \end{cases} \quad (8)$$

In our implementation, wavelet neurons are still injected to the reservoir in part, which is measured by the mixed ratio (MR) of injected wavelets, given by

$$MR = \frac{r_{wavelet}}{N} \quad (9)$$

where $r_{wavelet} \in \mathbb{Z}$ is the amount of injected wavelet neurons, N is the sum of all neurons. Besides, the dilation parameter a and translation parameter b are changed into decimals in order to balance the diversity of generated wavelet function family, given by

$$a_i = \frac{i}{r_{wavelet}},\ b_i = \frac{i}{r_{wavelet}} - 0.5,\ (i = 1, 2, ..., r_{wavelet}) \quad (10)$$

where b_i makes the center of each wavelet function family steady near the zero value to ensure the coordination between wavelet neurons and sigmoid neurons. Consequently, the activation functions based on SW, FBW and IRW are rewritten as follows

$$\psi^{SW}_{a_{i,\,b_i}}(t) = 2sinc(2(a_it - b_i) - 1) - sinc(a_it - b_i) \quad (11)$$

$$\psi^{FBW}_{a_{i,\,b_i}}(t) = C_{m,\,f_b} sinc^m(\frac{f_b(a_it - b_i)}{m}) e^{j2\pi f_c(a_it - b_i)} \quad (12)$$

$$\psi^{IRW}_{a_{i,\,b_i}}(t) = Ae^{-\frac{\beta}{\sqrt{1-\beta^2}}\omega_c(a_it - b_i)} sin(\omega_c(a_it - b_i)) \quad (13)$$

C. Network training

Similar to the typical ESN paradigm, the ESWN training is also to obtain the optimal output weight vector W^{out}. The training process includes the following stages.

(1) Network initialization

As discussed previously, elements of W and W^{in} are generated randomly during the initialization with random values drawn from a uniform distribution, and fixed until the end of ESWN execution. To account for echo state performance, the internal weight matrix W is typically scaled as

$$W \leftarrow \frac{\alpha}{|\lambda_{max}|} W \quad (14)$$

where $|\lambda_{max}|$ is the spectral radius of W, and $\alpha \in (0, 1)$ is a scaling parameter.

(2) Constructing wavelet-sigmoid reservoir

Especially, wavelet-neurons proportionally replace some original sigmoid-ones in a random way. Actually, the wavelet functions are embedded into corresponding neurons.

(3) Collecting network states

Network states before a washout time n_{min} is ignored due to the dependency on the initial states. The reservoir states obtained by Eq. (6) are collected in a state matrix X.

$$X = \begin{bmatrix} x^T(n_{min}+1) \\ x^T(n_{min}+2) \\ \vdots \\ x^T(n) \end{bmatrix} \quad (15)$$

and the corresponding target outputs are collected in a target output matrix Y

$$Y = \begin{bmatrix} y(n_{min}+1) \\ y(n_{min}+2) \\ \vdots \\ y(n) \end{bmatrix} \quad (16)$$

(4) Computing readout matrix

The readout matrix W^{out} is obtained by solving a linear regression problem, expressed as

$$XW^{out} = Y \quad (17)$$

Generally, the traditional method is to use the least squares solution, that is

$$W^{out} = \arg\min_w \|Xw - Y\| \quad (18)$$

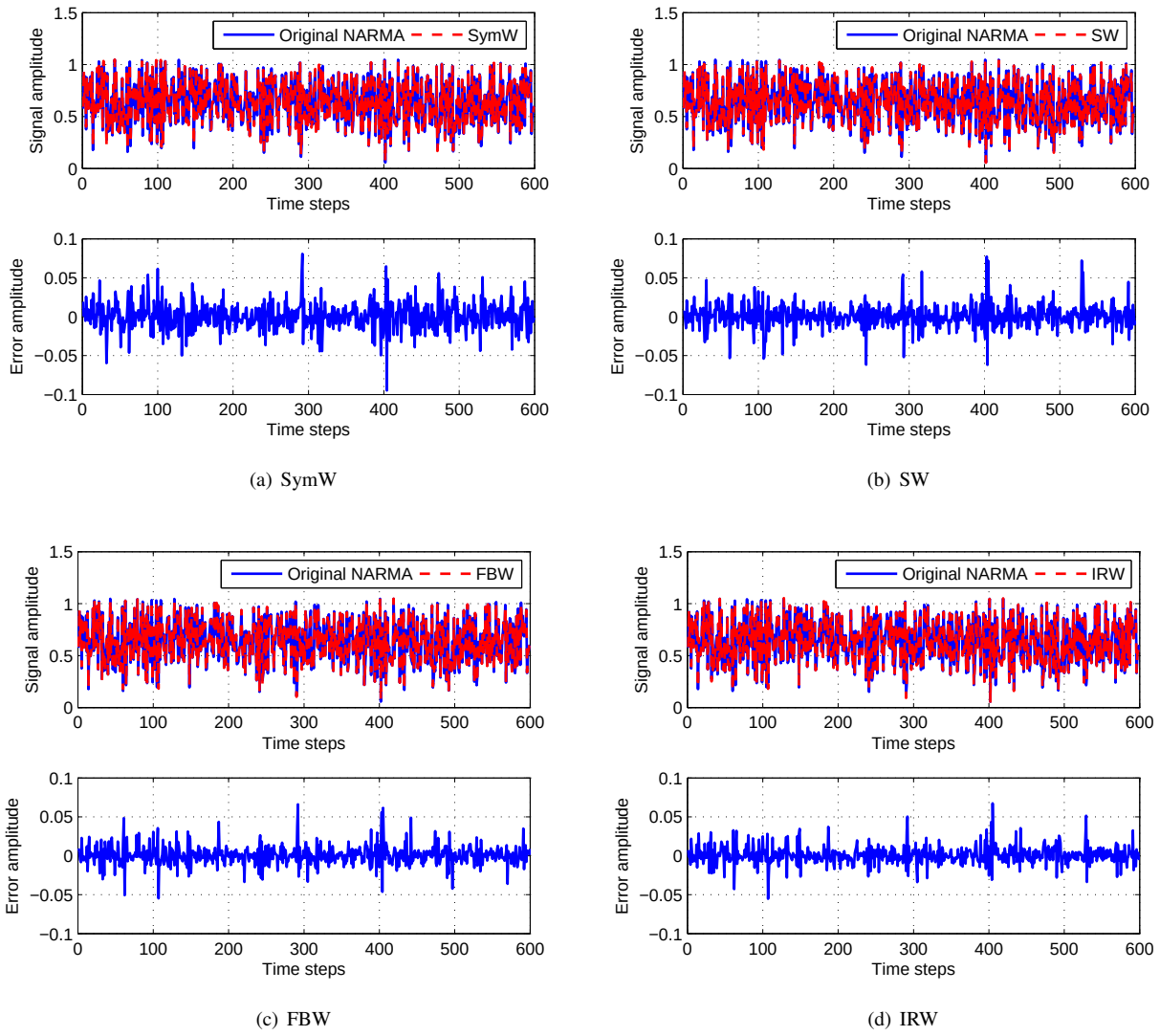
where $\|\cdot\|$ denotes the Euclidean norm. The output matrix W^{out} can be solved in a single step using the Moore-Penrose pseudo inverse

$$W^{out} = \tilde{X}Y = (X^TX)^{-1}X^TY \quad (19)$$

where $\tilde{X}$ denotes the generalized inverse of X. It means the completion of the local fine-tuning in regression layer. Afterwards, our ESWNs are used to perform nonlinear system identification tasks.

TABLE I
CONFIGURATION OF THE EVALUATED MODELS IN OUR EXPERIMENTS.

Parameters	NARMA	Ikeda map	MSO
Reservoir size	100	150	100
Spectral radius	0.7	0.7	0.6
MR	0.5	0.7	0.2
Reservoir Connectivity	0.1	0.15	0.2
Feedback weights between	[-0.1, 0.1]	[-0.56, 0.56]	[-0.25, 0.25]
Sequence length L	1400	5000	3000
L_{train}	700	2500	1500
L_{test}	700	2500	1500
Washout time L_v	100	100	100



(a) SymW (b) SW

(c) FBW (d) IRW

Fig. 2. Output and error of SymW, SW, FBW and IRW in the NARMA system.

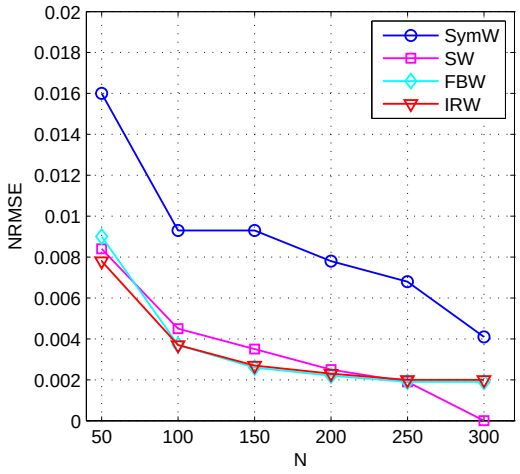


Fig. 3. Dependency of the network performance on the reservoir size for the NARMA system, utilizing SymW, SW, FBW and IRW.

III. Experiments

In the following section, we provide a comprehensive experimental evaluation of the ESWN models, considering three classical system identification tasks. In the proposed ESWNs, we build three novel reservoirs with wavelet- sigmoid neurons, using the Shannon wavelet, frequency B-spline wavelet and impulse response wavelet. For convenience of discussion, these models are called SW, FBW and IRW, respectively, We also evaluate the Symlets wavelet based ESWN, namely SymW, in order to demonstrate the advantages of our models. Table I shows the details of parameter configuration. The normalized root mean square error (NRMSE) is used to measure the nonlinear approximation capacity of the proposed models, defined as

$$NRMSE = \sqrt{\frac{\sum_{t=1}^{l_{test}} \left(y_{test}(t) - d(t)\right)^2}{l_{test} \cdot \sigma^2}} \tag{20}$$

where l_{test} is the length of test samples, $y_{test}(t)$ and $d(t)$ are the test output and desired output at time step t, respectively, and σ^2 is the variance of desired output $d(t)$. All the simulations are carried out in an identical software and hardware environment, and for each task, the average results over 10 trials are obtained for comparison.

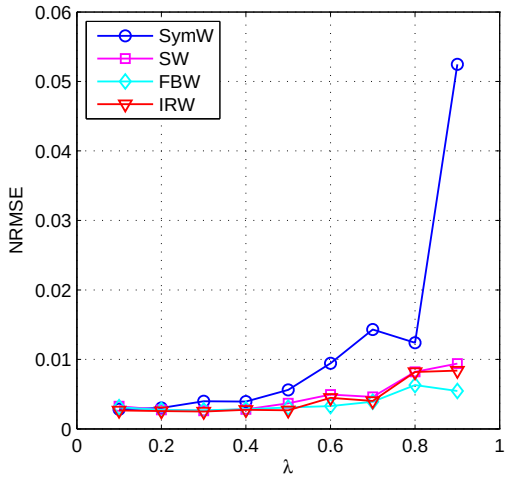


Fig. 4. Dependency of the network performance on the spectral radius for the NARMA system, utilizing SymW, SW, FBW and IRW.

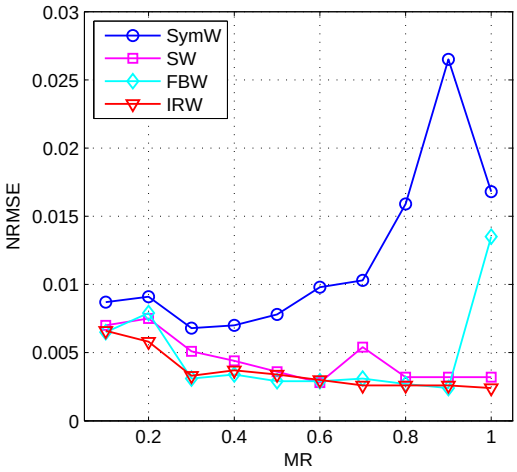
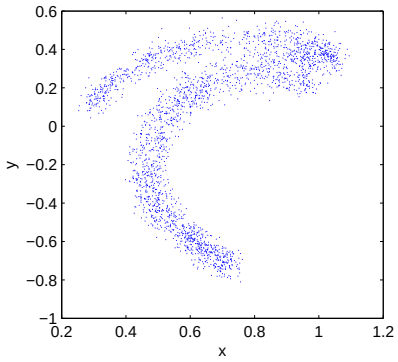
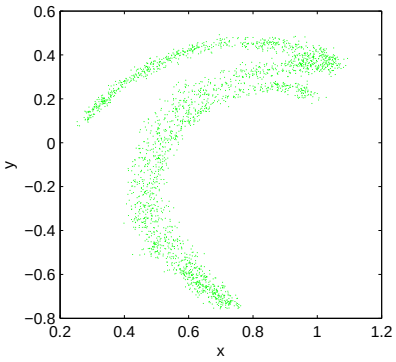


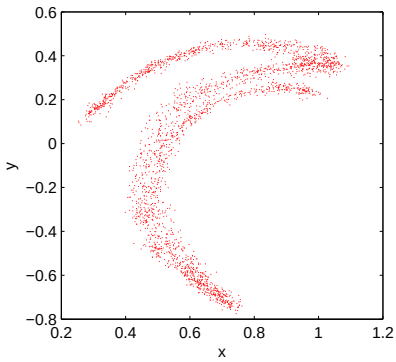
Fig. 5. Dependency of the network performance on the mixed ratio for the NARMA system, utilizing SymW, SW, FBW and IRW.



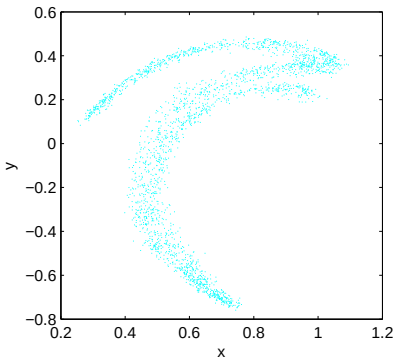
(a) Original Ikeda map



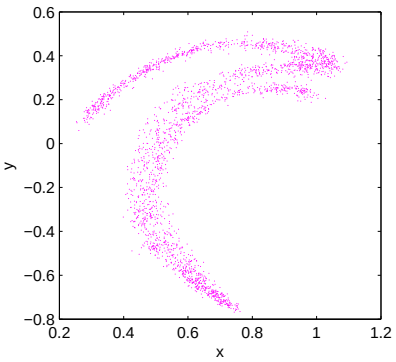
(b) SymW



(c) SW



(d) FBW



(e) IRW

Fig. 6. Attractors obtained by SymW, SW, FBW and IRW for the Ikeda map.

A. NARMA

The nonlinear autoregressive moving average (NARMA) system [30] is a discrete time system, whose current output depends on both the input and the previous output, given by

$$y(n) = 0.7x(n-m) + (1 - y(n-1))y(n-1) + 0.1 \quad (21)$$

where $y(n)$ is the system output at time n, and $x(t)$ is the system input at time n, m is the memory length (in my case, $m = 1$). In general, it is really difficult to model this system, due to the nonlinearity and possibly long memory.

Fig. 2 shows the comparative curves of the desired output and the estimated output as well as the output errors over time steps, for SymW, SW, FBW and IRW. As we observe from this figure, our models achieve more highly accurate approximations than SymW for the NARMA system. The corresponding NRMSEs are 0.0093, 0.0045, 0.0035 and 0.0037, respectively.

Figs. 3-5 show the average test set NRMSEs achieved by the selected method representatives versus different parameters for the NARMA system. In Fig. 3, we presents the dependency of network performance on the chosen reservoir size N. As we observe, SW, FBW and IRW achieve better

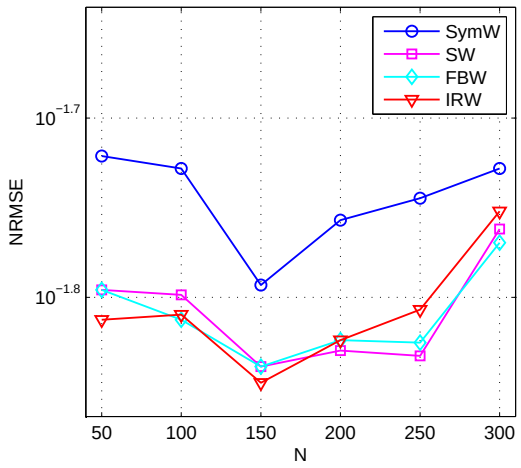


Fig. 7. Dependency of the network performance on the reservoir size for the Ikeda map, utilizing SymW, SW, FBW and IRW.

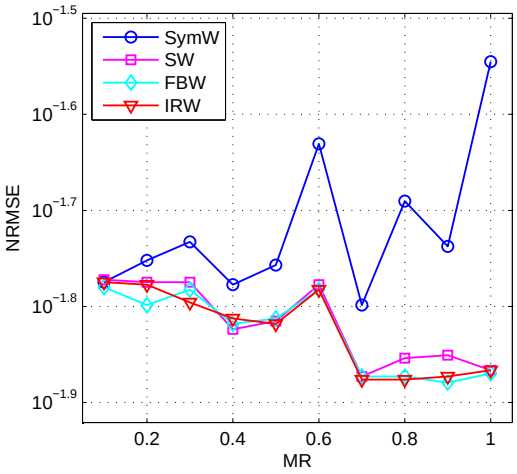


Fig. 9. Dependency of the network performance on the mixed ratio for the Ikeda map, utilizing SymW, SW, FBW and IRW.

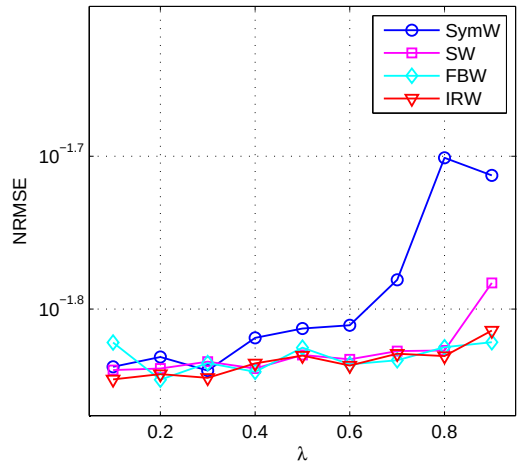


Fig. 8. Dependency of the network performance on the spectral radius for the Ikeda map, utilizing SymW, SW, FBW and IRW.

approximation than SymW, and the best scenario is that the reservoir size is $N = 300$. Fig. 4 shows the dependency of network performance on the spectral radius λ, where the proposed models obviously outperform SymW, and the relatively smaller λ is more conducive to superior performance. Furthermore, the dependency of network performance on mixed ratio is shown in Fig. 5. Likewise, the alternative models are superior to SymW over the whole choice of MR. It is noteworthy that their prediction performances start to get better as MR increases, while the performance of SymW tends to be worse.

B. Chaotic dynamical system

The Ikeda map [14],first proposed by Ikeda *et al.* (1980), is a discrete-time chaotic dynamical system, originated by the nonlinear optical system. As a model of laser light emission from a ring cavity containing a bistable dielectric medium, its 2D real example is given by the following form:

$$\begin{cases} x(n+1) = 1 + u(x(n)\sin t(n) - y(n)\cos t(n)) \\ y(n+1) = u(x(n)\sin t(n) + y(n)\cos t(n)) \end{cases} \quad (22)$$

where $x(n)$ and $y(n)$ denote the input and output of the Ikeda system at time n, the parameter u (in the case, $u = 0.7$) ensures that the system has a chaotic attractor, $t(n)$ is a time variable associated with $y(n)$ and $x(n)$, given by

$$t(n) = 0.4 - \frac{6}{1 + x^2(n) + y^2(n)} \quad (23)$$

where the system input and output are initialized to $x(0) = 0.1$ and $y(0) = 0.1$. And there exists the gaussian white noise with the standard deviation $\upsilon \in (0, 1)$ in the system.

In Fig. 6, we show the chaotic attractor of the original Ikeda map and ones produced by the evaluated models. As seen from this figure, all the trained models come close to the trajectory of the original Ikeda map attractor. The corresponding prediction NRMSEs are 0.0160, 0.0146, 0.0145 and 0.0142, respectively.

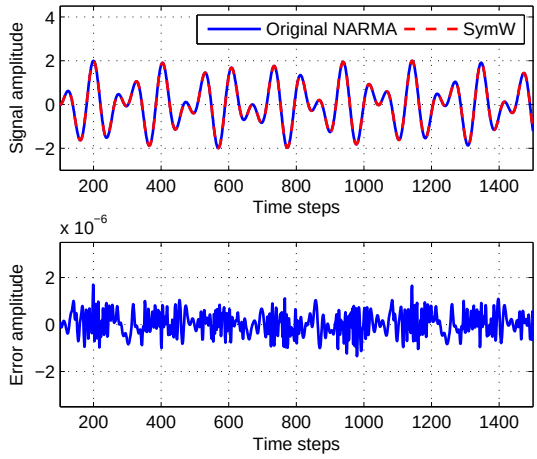
Figs. 7-9 shows the average test set NRMSEs achieved by the selected model representatives versus different parameters for the Ikeda map system. In all cases where the considered reservoir size changes, the proposed models can perform better than SymW, as shown in Fig. 7. The optimal case is that the reservoir size is set to 150. Likewise in Figs. 8-9, our models are also superior to SymW over the choices of λ and MR. Specially, SW, FBW and IRW are slightly sensitive to the change of λ, but the approximation performance of SymW degenerates gradually with the increase of λ. In addition, the larger MR is considered, the better approximations are obtained for our models, and MR should be kept up within the set range [0.7, 1] to ensure the satisfactory approximation performance as much as possible in the Ikeda map.

C. Multiple superimposed oscillations

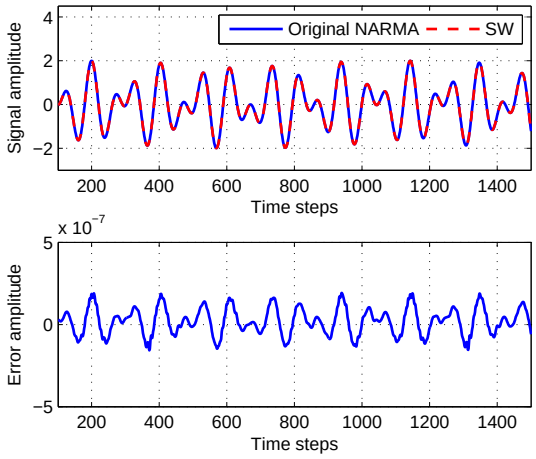
The considered MSO system [12] are built by summing up several simple sine wave functions. Formally the general expression is given by the following equation

$$y(n) = \sum_{i=1}^{s} \sin(\alpha_i n) \quad (24)$$

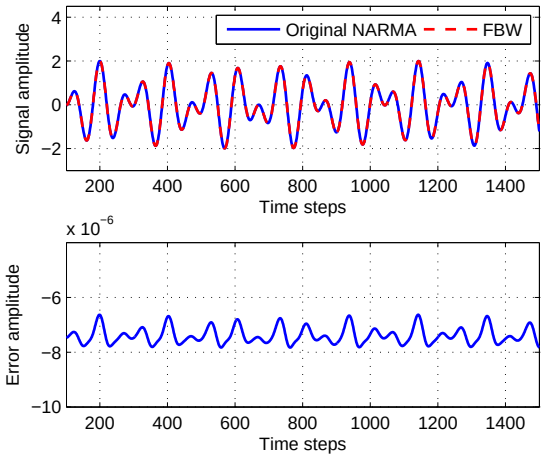
where s denotes the number of sine waves, α_i denotes the frequencies of the summed sine waves, and n specifies an



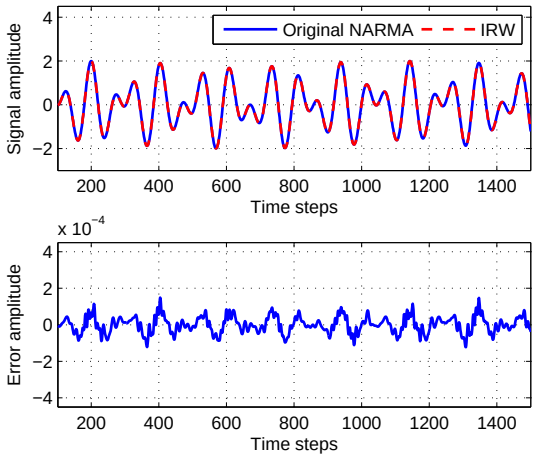
(a) SymW



(b) SW



(c) FBW



(d) IRW

Fig. 10. Output and error of SymW, SW, FBW and IRW in the MSO system.

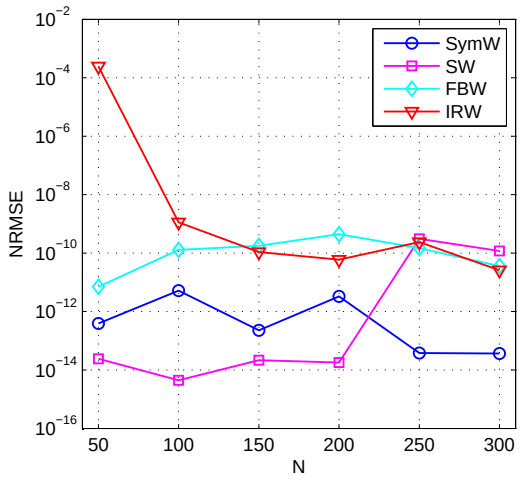


Fig. 11. Dependency of the network performance on the reservoir size for the MSO system, utilizing SymW, SW, FBW and IRW.

integer index of the time step. We use MSOx to describe the especial MSO dynamics, where x defines the number of summed sine waves. To demonstrate the capability of the ESWNs in approximating the MSO system, they are trained to learn the function composed of two sines, i.e., MSO2:

$$y(n) = sin(0.2n) + sin(0.311n) \quad n = 1, 2, ... \tag{25}$$

Fig. 10 shows comparative curves of the desired network output and actual signal versus time steps as well as the corresponding output error for each model under consideration. It is observed that our models again demonstrate their abilities to identify the nonlinear system with high approximation performance. The obtained NRMSEs are 2.1×10^{-13}, 7×10^{-15}, 5.5×10^{-11} and 2.1×10^{-9}, respectively. Moreover, it can be clearly seen that the test NRMSE for SW is of two orders of magnitudes better than the NRMSE achieved by the other alternatives, which implies that it can track the oscillations more accurately.

In Fig. 11-13, we provide the average test set NRMSEs achieved by the selected model representatives versus different parameters for the MSO system. From the figures, we observe that SW is the best choice for identifying this system. In particular, the region where SW shows the best approximation lies within the range $N \in [50, 200]$ and $\lambda \in [0.6, 1]$ (see Figs 11-12). But for MR, SW works considerably better than the other evaluated models. Besides, all structures except IRW tend to perform poorly as MR increases up to 1 (see Fig. 13).

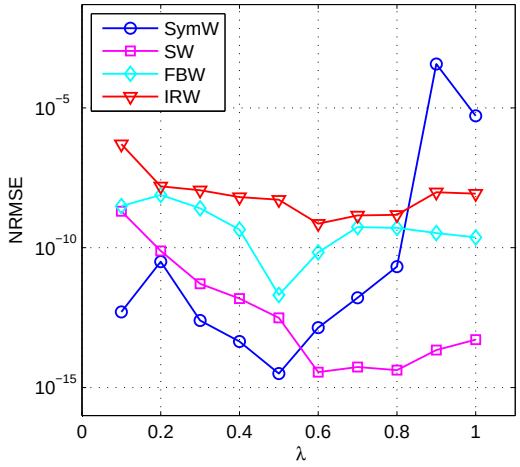


Fig. 12. Dependency of the network performance on the spectral radius for the MSO system, utilizing SymW, SW, FBW and IRW.

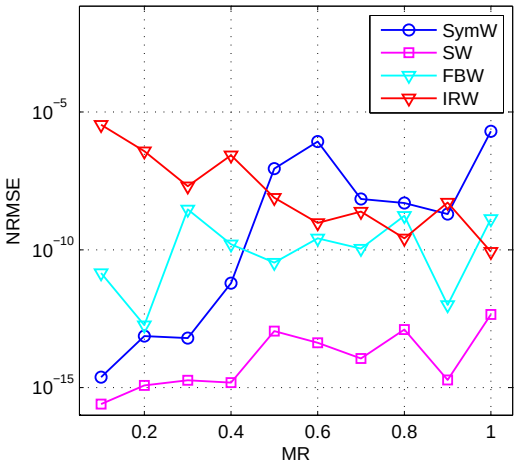
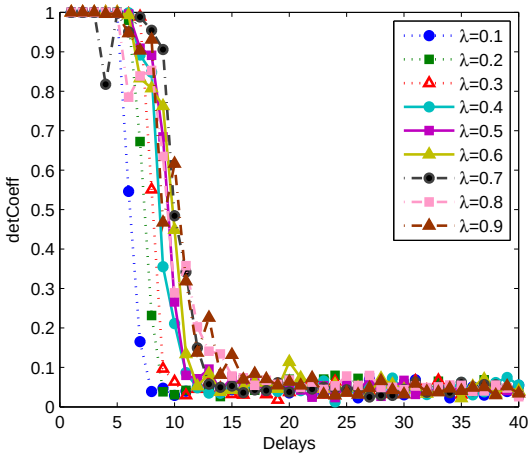
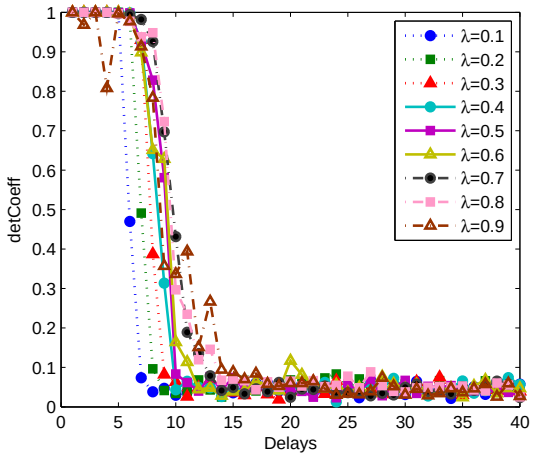


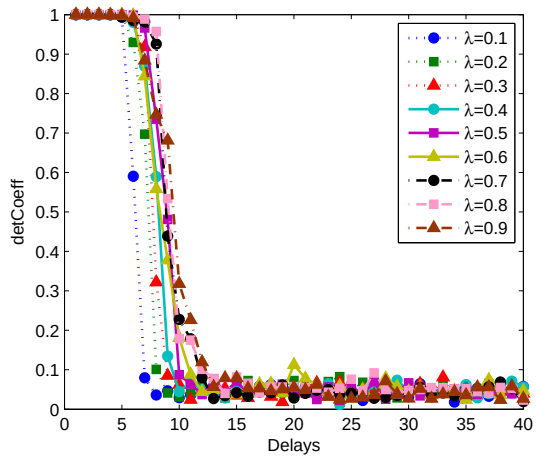
Fig. 13. Dependency of the network performance on the mixed ratio for the MSO system, utilizing SymW, SW, FBW and IRW.



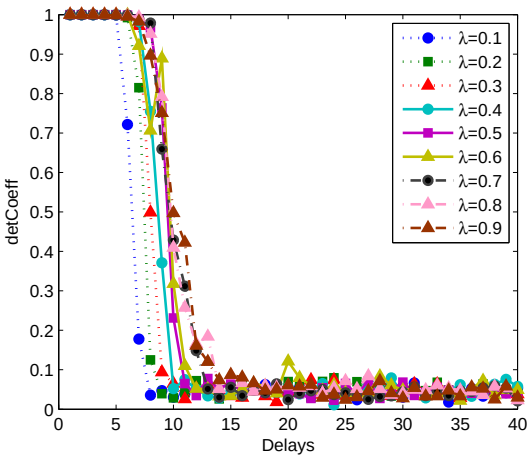
(a) SymW



(b) SW



(c) FBW



(d) IRW

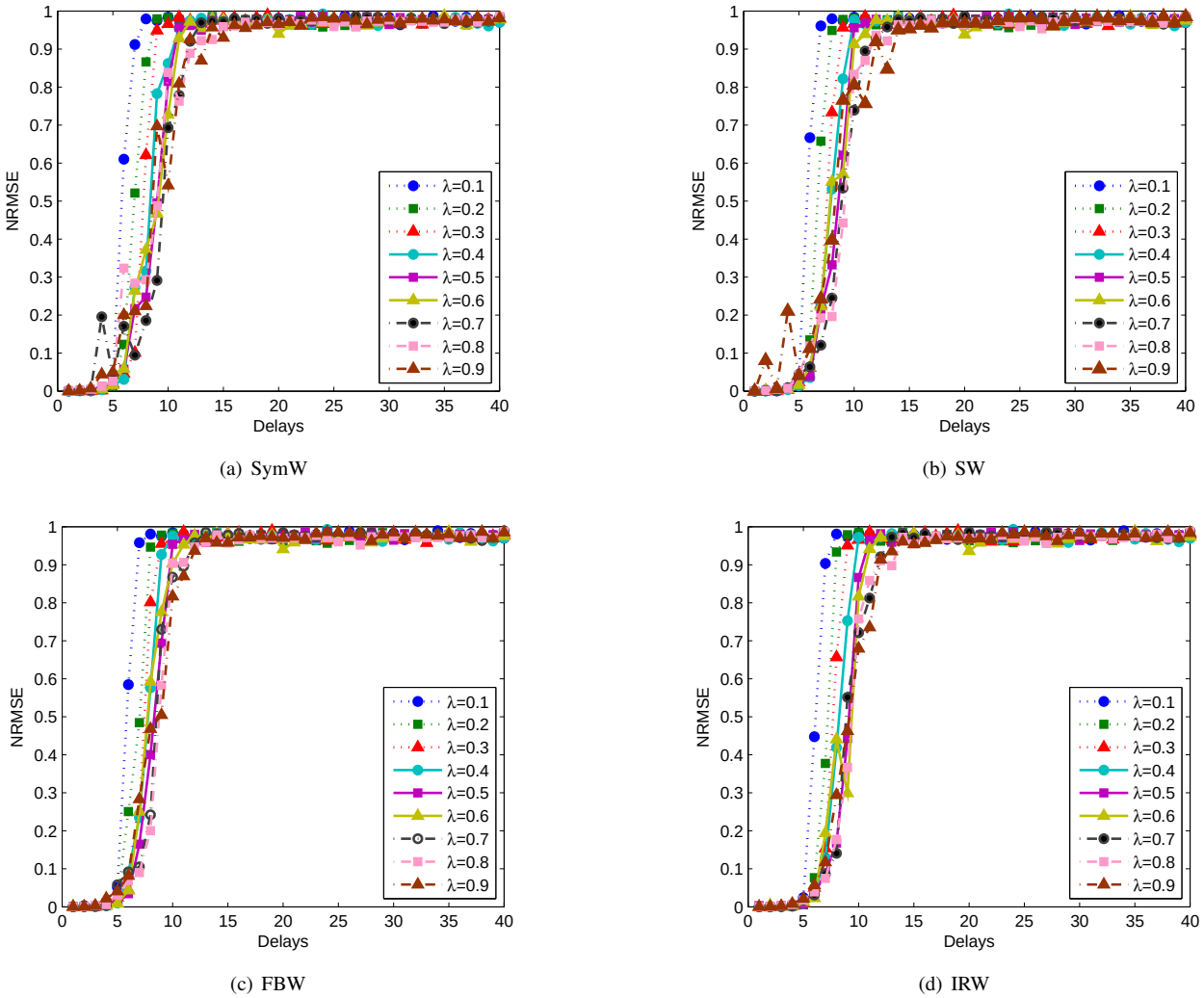
Fig. 14. The forgetting curves of the evaluated models versus different λ.

IV. Short-term memory capacity

In [32], the inherent capacity of ESN reservoir architectures is quantified by representing past events through a measure correlating the past events in an $i.i.d.$ input stream with the network output. As an outstanding feature of reservoir, the memory capacity (MC) can embody the property of some input-output systems. Just for the sake of argument, assume that the network is driven by a univariate stationary input signal $u(t)$. For a given delay k, the network with optimal parameters is established for the task of outputting $u(t-k)$

TABLE II
STM CAPACITIES OF SYMW, SW, FBW AND IRW VERSUS THE SPECTRAL RADIUS λ.

Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
SymW	7.15	8.5318	9.0226	9.7414	10.2162	10.5053	10.8641	10.6701	10.9794
SW	6.9878	8.1992	8.7930	9.3899	9.8149	9.9330	10.5855	10.8559	10.3364
FBW	7.1000	8.3710	8.7473	9.0656	9.6567	9.5561	9.9788	10.4384	10.2556
IRW	7.3346	8.5796	8.9694	9.6099	10.3618	10.4794	10.7050	11.1950	11.1458



(a) SymW (b) SW

(c) FBW (d) IRW

Fig. 15. Reconstruction NRMSEs of the evaluated models versus different λ.

in the case that the $i.d.d.$ input stream $...u(t-2)u(t-1)u(t)$ is obtained up to time t. The capacity to recover the values of former inputs is measured in terms of the squared correlation coefficient between the desired output (input signal delayed by k time steps) and the observed network output $y(t)$.

$$MC_k = \frac{Cov^2(u(t-k), y(t))}{Var(u(t))Var(y(t))} \tag{26}$$

where Cov and Var denote the covariance and variance, respectively. The short-term memory (STM) capacity is then formulated as:

$$MC = \sum_{k=1}^{\infty} MC_k \tag{27}$$

And in [32], Jaeger proved that the STM capacity cannot surpass the reservoir size N under the assumption of $i.i.d.$ input stream.

We empirically evaluate the short-term MC of the considered models, where they are trained to memorize the inputs delayed by $k = 1, 2, \ldots, 40$. Consider that these models are equipped with 1 input unit, 20 wavelet-sigmoid reservoir units and 40 output units. The input signal subjects to random uniform distribution in the interval $[0, 0.5]$, and MR is set to 0.2. The desired output is a series of delayed versions of the input signal [11][14][32].

Fig. 14 shows the forgetting curves of the evaluated models over a wide choice of λ, where the $detCoeff$ is the squared

TABLE III
RECONSTRUCTION PERFORMANCE OF SYMW, SW, FBW AND IRW VERSUS SPECTRAL RADIUS λ.

Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
SymW	0.8450	0.8164	0.8000	0.7881	0.7771	0.7758	0.7674	0.7753	0.7704
SW	0.8475	0.8224	0.8065	0.7959	0.7863	0.7875	0.7731	0.7697	0.7862
FBW	0.8460	0.8206	0.8080	0.8012	0.7882	0.7933	0.7832	0.7735	0.7835
IRW	0.8404	0.8131	0.8018	0.7887	0.7716	0.7730	0.7662	0.7586	0.7631

correlation coefficient (i.e., MC_k in (26)), and the corresponding MC values are depicted in Table II. It can be seen clearly that IRW has a superior STM compared with the other models, and interestingly, their STMs become more powerful as λ increases. To illustrate the approximation capacity of ESWNs, we show the NRMSE of each reconstruction in the generating Fig. 15, and calculate the average NRMSE across 40 reconstructions for each model, as shown in Table III. The experimental results show that IRW has a superior ability of recovering the network inputs with high performance. For example, for $\lambda = 0.8$, the reconstruction performance of IRW is increased to around 69%, which is 1.6%, 1% and 1.5% higher than those obtained by SymW, SW and FBW, respectively. More importantly, combined with Table II, it can be seen that the more powerful the STM capacity is, the better nonlinear approximation capacities ESWNs have.

V. DISCUSSION

The paper focuses on finding more outstanding wavelets instead of the common Symlets wavelet towards the hybrid reservoir to achieve superior nonlinear approximation capacity. Three ESWNs are proposed, considering Shannon wavelet, frequency B-spline wavelet and impulse response wavelet. On the one hand, the transformation of a homogeneous reservoir into a novel hybrid wavelet-sigmoid one allows a more effective states representation. On the other hand, the harmonious cooperation between the new wavelet neurons and sigmoid neurons offers us a novel nonlinear approximation capability.

In all the computer experiments conducted in this paper, we have demonstrated that the proposed ESWNs, i.e., SW, FBW or IRW, possess the more powerful nonlinear approximation capacity, compared with the common SymW (see Figs 2-13). Furthermore, we choose the representative parameters to confirm the effectiveness of ESWNs for the given tasks, where the parameters are specified by reservoir size, spectral radius, and mixed ratio of wavelets. The reservoir size determines ESN's nonlinear mapping capacity. Generally speaking, the bigger reservoir size is beneficial to obtain the better approximation performance, but the appropriate regularization should be taken against overfitting. The spectral radius measures how fast the influence of an input fades away in a reservoir over time, and whether the reservoir is stable or not. As previously discussed in [33], the larger spectral radius tends to obtain the slower decay of the networks response to an impulse input, and the stronger networks memory capacity. In this case, ESN can have a more efficient computing power and a better approximation capability. The mixed ratio of wavelet neurons determines the feature of reservoir (unique sigmoid reservoir or hybrid wavelet-sigmoid reserovir), as well as the activation level of reservoir neurons. Unfortunately, this parameter has been never deeply discussed in existing literatures [24][25]. In fact, these reservoir parameters are task-dependent, as shown in Figs. 3-5, 7-9 and 11-13. For example, considering the mixed ratio of wavelet neurons, we conclude that IRW and FBW seem to be suitable for the NARMA and Ikeda map tasks (see Fig. 5 and Fig. 9), while SW is the optimal selection for the MSO task (see Fig. 13). Moreover, for the NARMA and Ikeda map tasks, the approximation performance of our models is directly proportional to the mixed ratio of wavelet neurons, but the contrary is the case for the MSO task. The determination of these reservoir parameters for a given task is still an open problem, and it is worthy of further research.

VI. CONCLUSION

In this paper, three echo state wavelet-sigmoid networks are proposed considering Shannon wavelet, frequency B-spline wavelet and impulse response wavelet. We have focused on the following research issues.

(1) Whether or not the alternative wavelets are suitable for reservoir activations so that the constructed ESWNs are comparable to the Symlet wavelet based one, and even outperform it?
(2) What kind of reservoirs are required to construct competitive ESWNs?
(3) When competitive wavelet-sigmoid reservoirs are constructed, how do they compare in terms of MC with established one (built-in Symlets wavelet), and what is the relationship between MC and NRMSE?

Through three widely used system identification benchmarks of different origins and characteristics, as well as by performing a theoretical analysis, we have demonstrated the following.

(1) The proposed SW, FBW and IRW can achieve more highly accurate approximation compared with SymW.
(2) The reservoirs can be constructed by means of introducing the wavelet-sigmoid neurons.
(3) The MCs of our models are directly proportional to their reconstruction NRMSE, and IRW has the most powerful MC and nonlinear approximation capacity.

Selecting a suitable activation function is crucial to the accuracy and computational efficiency of ESNs. Future research will work on optimizing activation function to address the difficulty in selecting suitable activation function or combination of functions for a given task.

REFERENCES

[1] F. J. Ordonezw, D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," Sensors, vol. 16, no. 1, pp. 1-25, 2016.

[2] L. Yang, Y. Yang, Y. Li, T. Zhang, "Almost Periodic Solution for a Lotka-Volterra Recurrent Neural Networks with Harvesting Terms on Time Scales," Engineering Letters, vol. 24, no. 4, pp. 455-460, 2016.

[3] J. Y. B. Tan, D. B. L. Bong, A. R. H. Rigit, "Time Series Prediction using Backpropagation Network Optimized by Hybrid K-means-Greedy Algorithm," Engineering Letters, vol. 20, no. 3, pp. 203-210, 2012.

[4] D. T. Mirikitani, N. Nikolaev, "Recursive bayesian recurrent neural networks for time-series modeling[J]. IEEE Transactions on Neural Networks," vol. 21, no. 2, pp. 262-274, 2010.

[5] M. Lukosevicius, H. Jaeger, B. Schrauwen, "Reservoir computing trends," KI-Knstliche Intelligenz, vol. 26, no. 4, pp. 365-371, 2012.

[6] S. P. Chatzis, Y. Demiris, "Echo State Gaussian Process," IEEE Transactions on Neural Networks, vol. 22, no. 9, pp. 1435-1445, 2011.

[7] M. C. Ozturk, J. C. Principe, "An associative memory readout for ESNs with applications to dynamical pattern recognition," Neural Networks, vol. 20, no. 3, pp. 377-390, 2007.

[8] S. I. Han, J. M. Lee, "Precise Positioning of Nonsmooth Dynamic Systems Using Fuzzy Wavelet Echo State Networks and Dynamic Surface Sliding Mode Control," IEEE Transactions on Industrial Electronics, vol. 60, no. 11, pp. 5124-5136, 2013.

[9] Z. K. Malik, A. Hussain, J. Wu, "Novel Biologically Inspired Approaches to Extracting Online Information from Temporal Data," Cognitive Computation, vol. 6, no. 3, pp. 1-13, 2014.

[10] C. Sheng, J. Zhao, Y. Liu, W. Wang, "Prediction for noisy nonlinear time series by echo state network based on dual estimation," Neurocomputing, vol. 82, no. 4, pp. 186-195, 2012.

[11] A. Rodan, P. Tino, "Minimum complexity echo state network," IEEE Transactions on Neural Networks, vol. 22, no. 1, pp. 131-144, 2011.

[12] D. Koryakin, J. Lohmann, M. V. Butz, "Balanced echo state networks," Neural Networks, vol. 36, no. 2012, pp. 35-45, 2012.

[13] C. Gallicchio, A. Micheli, "Tree Echo State Networks," Neurocomputing, vol. 101, no. 3, pp. 319-337, 2013.

[14] X. C. Sun, H. Y. Cui, R. P. Liu, J. Y. Chen, Y. J. Liu, "Modeling deterministic echo state network with loop reservoir," Frontiers of Information Technology & Electronic Engineering, vol. 13, no. 9, pp. 689-701, 2012.

[15] S. P. Chatzis, Y. Demiris, "The copula echo state network," Pattern Recognition, vol. 45, no. 1, pp. 570-577, 2012.

[16] Z. Xia, C. Yuan, X. Sun, D. Sun, R. Lv, "Combining Wavelet Transform and LBP Related Features for Fingerprint Liveness Detection," IAENG International Journal of Computer Science, vol. 43, no. 1, pp. 290-298, 2016.

[17] A. S. Muthanantha Murugavel, S. Ramakrishnan, "An Optimized Extreme Learning Machine for Epileptic Seizure Detection," IAENG International Journal of Computer Science, vol. 41, no. 4, pp. 212-221, 2014.

[18] A. Ismail, D. S. Jeng, L. L. Zhang, J. S. Zhang, "Predictions of bridge scour: application of a feed-forward neural network with an adaptive activation function," Engineering Applications of Artificial Intelligence, vol. 26, no. 5, pp. 1540-1549, 2013.

[19] R. X. Gao, R. Yan, "Wavelets: Theory and applications for manufacturing," Springer Science & Business Media, 2010.

[20] J. Cohen, A. I. Zayed, "Wavelets and multiscale analysis: theory and applications," Springer Science & Business Media, 2011.

[21] S. K. Jain, S. N. Singh, "Low-order dominant harmonic estimation using adaptive wavelet neural network," IEEE Transactions on Industrial Electronics, vol. 61, no.1, pp. 428-435, 2014.

[22] K. Bhaskar, S. N. Singh, "AWNN-assisted wind power forecasting using feed-forward neural network," IEEE transactions on sustainable energy, vol. 3, no. 2, pp. 306-315, 2012.

[23] Y. Bodyanskiy, O. Vynokurova, "Hybrid adaptive wavelet-neuro-fuzzy system for chaotic time series identification," Information Sciences, vol. 220, no. 2013, pp. 170-179, 2013.

[24] S. Wang, X. Yang, C. Wei, "Harnessing Non-linearity by Sigmoid-wavelet Hybrid Echo State Networks," The Sixth World Congress on Intelligent Control and Automation, IEEE, 2006.

[25] H. Cui, C. Feng, Y. Chai, R. P. Liu, Y. J. Liu, "Effect of hybrid circle reservoir injected with wavelet-neurons on performance of echo state network," Neural Networks, vol. 57, no. 9, pp. 141-151, 2014.

[26] J. Rafiee, M. A. Rafiee, P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," Expert Systems with Applications, vol. 37, no. 6, pp. 4568-4579, 2010.

[27] C. Cattani, "Fractional Calculus and Shannon Wavelet," Mathematical Problems in Engineering, vol. 2012, no. 2012, pp. 1-26, 2012.

[28] Z. Yang, X. Chen, X. Zhang, Z. He, "Free vibration and buckling analysis of plates using B-spline wavelet on the interval Mindlin element," Applied Mathematical Modelling, vol. 37, no. 5, pp. 3449-3466, 2013.

[29] L. Alkafafi, C. Hamm, T. Sauer, "Vibrational Error Extraction Method Based on Wavelet Technique," Mathematical Methods for Curves and Surfaces, Springer Berlin Heidelberg, 2012.

[30] H. Jaeger, "The echo state approach to analyzing and training neural networks," German Nat. Res. Inst. Inform. Technol., Sankt Augustin, Germany, Rep, 2002.

[31] G. Holzmann, H. Hauser, "Echo state networks with filter neurons and a delay&sum readout," Neural Networks, vol. 23, no. 2, pp. 244-256, 2010.

[32] H. Jaeger, "Short Term Memory in Echo State Networks," GMD Report, 2002.

[33] Z. Deng, Y. Zhang, "Collective behavior of a small-world recurrent neural system with scale-free distribution," IEEE Transactions on Neural Networks, vol. 18, no. 5, pp. 1364-1375, 2007.

Xiaochuan Sun received the M.S. degree from Guilin University of Electronic Technology, Guilin, China, in 2010, and the Ph.D. degree from Beijing University of Posts and Telecommunications, Beijing, China, 2013, respectively. He works at the College of Information Engineering, North China University of Science and Technology, Tangshan, and is currently a lecturer. The current research interests include complex network theory, reservoir computing, swarm intelligence theory.

Yingqi Li received the B.S. and M.S. degrees in Guilin University of Electronic Technology, Guilin, China, in 2007 and 2010, respectively. She has been with the College of Electrical Engineering, North China University of Science and Technology, Tangshan, since 2010, and is currently a lecturer. The current research interests include statistical pattern recognition, evolutionary computation, neural networks.