Jointly optimized echo state network for short-term channel state information prediction of fading channel

Qiwu Luo[1], *Member, IEEE*, Jian Zhou[1], Yichuang Sun[2], *Senior Member, IEEE*, Oluyomi Simpson[2] *Member, IEEE*
[1] School of Automation, Central South University, Changsha 410083, P. R. China
[2] School of Engineering and Computer Science, University of Hertfordshire, Hatfield, Herts AL10 9AB, United Kingdom.
E-mail: luoqiwu@csu.edu.cn

***Abstract*—Accurately obtaining channel state information (CSI) in wireless systems is significant but challenging. This paper focuses the technique of machine-learning-based channel estimation. In particular, a jointly optimized echo state network (JOESN) is proposed to form a concept of the CSI prediction which is made up of two interacting aspects of output weight regularization and initial parameter optimization. First, in order to enhance noise robustness, a sparse regression based on L_2 regularization is employed to finely learn the output weights of ESN. Second, vital reservoir parameters (i.e., global scaling factor, reservoir size, scaling coefficient and sparsity degree) are learned by a linear-weighted particle swarm optimization (LW-PSO) for further improve the prediction accuracy and reliability. The experiments about computational complexity and three evaluating metrics are carried out on two chaotic benchmarks and one real-world dataset. The analyzed results indicate that the JOESN performs promisingly on multivariate chaotic time series prediction.**

Keywords—Wireless communications, channel state information (CSI), channel estimation, fading channel, reservoir computing.

I. Introduction

In wireless communication systems, transmitted signal is inevitably suffering with degradation during the propagation process due to stochastic intercarrier interference and time-varying channel fading [1]. In order to effectively overcome the interference and distortion, the receiver usually needs to obtain accurate channel state information (CSI). Consequently, efficient and precise channel estimation plays a core role to sustain wireless communication system be in stably health status. Generally, the CSI can be obtained either through a blind method utilizing the statistical knowledge of the channel or a training-based method based on priori known pilot symbols.[2]. However, the blind estimation techniques can be only applied to the relatively-slow-varying channel, while the training-based methods are reported to be inefficient for requiring dedicated time-frequency resources [3].

In recent years, stimulated by many applications in artificial intelligence fields, a special group of training-based, namely machine-learning-based methods, are dedicated to deal with problems in wireless communication area by establishing a straightforward relation between the input and output of the system [4]. Inspired by this research trend, this paper makes attempt to find a potential solution to actively learn (or say predict) the time-varying rules of the CSI, comparing with the passive channel estimation.

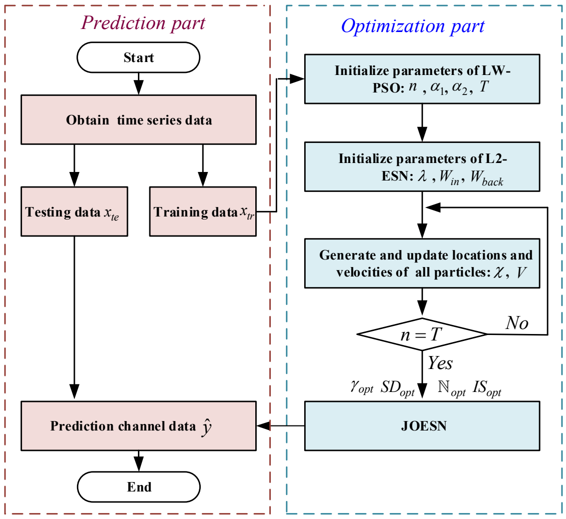

Fig. 1. The flow chart of JOESN.

As one of the most powerful reservoir computing scheme, the echo state network (ESN) performs much better than many other recurrent neural networks on computational expense and convergence speed, which has been innovatively used to accurately obtain CSI in MIMO-OFDM systems [5]. Notably, ESN has exhibited evident energy efficiency in next generation of green communications [3]. Imagine that ESN- based symbol detection is accepted in wireless communications, the tough task of channel estimation can be implicitly achieved by the CSI prediction.

However, several crucial constraints such as collinearity problems limit the extensive applications of ESN, especially when data incorporate random noise, but data collected from actual wireless channel are always noisy. In this paper, a jointly optimized echo state network (JOESN) is proposed, where the fundamental ESN (FESN) is optimized from two aspects. (1) Outlet side: output weight regularization, and (2) Internal side: reservoir parameter optimization and status supervision. The main contributions of this work are summarized as follows.

1) A sparse regression based on L_2 regularization is employed to avoid ill-posed solutions. Then the sufficiently unbiased and small-amplitude output weights can be obtained in the L_2-ESN version.

2) A linear-weighted particle swarm optimization (LW-PSO) is adopted in the L_2-ESN to optimize network parameters intelligently and supervise the reservoir status implicitly.

Hence, the L_2-ESN optimized by LW-PSO (namely, JOESN) can escape from inefficient and unreliable manual settings.

3) Extensive evaluations (i.e., computational complexity, root mean square error (RMSE), normalized RMSE (NRMSE), symmetric mean absolute percentage error (SMAPE)) are carried out and discussed on two chaotic benchmarks and one dataset of Rayleigh fading channel.

The structure of this paper is as follows: In Section II, the basic theories of ESN and PSO are briefly reviewed, then we introduce the proposed JOESN in detail, and its computation complexity is analyzed in Section III. Section IV evaluates the performance of JOESN on two chaotic systems. And Section V illustrates and discusses the application performance of ESN on Rayleigh fading channel. Section VI concludes the paper.

II. JOINTLY OPTIMIZED ECHO STATE NETWORK

A. Echo State Network with L_2 Regularization

ESN was first proposed by Jaeger and Hass [6] for predicting chaotic systems. A typical ESN includes three layers, input layer, hidden layer (reservoir) and output layer. The state variables can be formulated as

$$\begin{cases} \boldsymbol{U}(t)=[u_1(t),u_2(t),...,u_M(t)]^T \\ \boldsymbol{X}(t)=[x_1(t),x_2(t),...,x_N(t)]^T \\ \boldsymbol{Y}(t)=[y_1(t),y_2(t),...,y_L(t)]^T \end{cases} \tag{1}$$

where $\boldsymbol{U}\in\mathbb{R}^{M\times L}$, $\boldsymbol{X}\in\mathbb{R}^{N\times L}$ and $\boldsymbol{Y}\in\mathbb{R}^{L\times 1}$ are input discrete time signals, internal neurons and output signals, respectively. And *M, N* and *L* are the dimension of the input layer, internal reservoir and output layer, respectively.

The internal reservoir is updated by

$$\boldsymbol{X}=\boldsymbol{\Gamma}(\boldsymbol{W}_{in}\boldsymbol{U}+\boldsymbol{W}\boldsymbol{X}+\boldsymbol{W}_{back}^T\boldsymbol{Y}^T) \tag{2}$$

and the output data is obtained by

$$\boldsymbol{Y}=[\boldsymbol{U}:\boldsymbol{X}]^T\boldsymbol{W}_{out} \tag{3}$$

where $\boldsymbol{\Gamma}(\bullet)$ and $[\bullet:\bullet]$ are the active function and matrix connection, respectively. It is noted that $\boldsymbol{\Gamma}(\bullet)$ is often defined as the hyperbolic tangent function (tanh). And $\boldsymbol{W}_{in}\in\mathbb{R}^{N\times M}$, $\boldsymbol{W}\in\mathbb{R}^{N\times N}$, $\boldsymbol{W}_{out}\in\mathbb{R}^{(N+M)\times 1}$, and $\boldsymbol{W}_{back}\in\mathbb{R}^{1\times N}$ are input weights, updated weights, output weights and echo weights, respectively. The output weights $\boldsymbol{W}_{out}$ are calculated by minimizing the following square error

$$E_{org}=\min\left\|\boldsymbol{Y}-[\boldsymbol{U}:\boldsymbol{X}]^T\boldsymbol{W}_{out}\right\|_2^2 \tag{4}$$

A sparse regression based on L_2 regularization is employed to finely learn the output weights. Then the loss function in (4) is modified as

$$E_{regu}=\min\left\{\left\|\boldsymbol{Y}-[\boldsymbol{U}:\boldsymbol{X}]^T\boldsymbol{W}_{out}\right\|_2^2+\lambda\left\|\boldsymbol{W}_{out}\right\|_2^2\right\} \tag{5}$$

where $\lambda\in(0,1)$ is the penalty parameter of L_2 regularization. Hence the estimated output weight $\boldsymbol{W}_{out}$ is calculated by

$$\boldsymbol{W}_{out}=([\boldsymbol{U}:\boldsymbol{X}][\boldsymbol{U}:\boldsymbol{X}]^T+\lambda\boldsymbol{E})^{-1}[\boldsymbol{U}:\boldsymbol{X}]\boldsymbol{Y} \tag{6}$$

where $\boldsymbol{E}$ is an identity matrix. There are four key parameters in the dynamic reservoir, jointly determining the final performance of ESN, which are spectral radius (*SR*), reservoir size ($N\in Z^+$), input layer scaling coefficient ($IS\in(0, 1]$), and reservoir sparsity degree ($SD\in(0, 1)$). Finally, the output are estimated as $[\boldsymbol{U}:\boldsymbol{X}]\boldsymbol{W}_{out}$.

B. Linear-Weighted Particle Swarm Optimization

The particle swarm optimization (PSO) is very specialized in searching fitting solutions for complex temporal problems. A standard PSO is expressed as

$$V_i(t+1)=\varpi V_i(t)+\alpha_1 r_1(P_{i,d}-\chi_i(t))+\alpha_2 r_2(P_g-\chi_i(t)) \tag{7}$$

$$\chi_i(t+1)=\chi_i(t)+V_i(t+1) \tag{8}$$

where t and i are the iteration index and particle index, $V_i(t+1)$and $\chi_i(t+1)$ denote the flight velocity and spatial vector of the i-th particle in (t+1)-th iteration, ϖ is the inertia factor, $\alpha_1$and α_2 are the acceleration parameters, r_1 and r_2 are two random values belonging to (0,1). $P_{i,d}$ and P_g record the spatial vector with the optimal solution of the current particle and all particles, respectively.

A linear-weighting scheme is imported to prevent PSO from ripping into local optimal locations. Then a well global search ability at the early optimization stage and a high local search performance at the later stage can be obtained simultaneously. This linear weight parameter is expressed as

$$\varpi=\varpi_{\max}-\frac{t}{T}(\varpi_{\max}-\varpi_{\min}) \tag{9}$$

where $\varpi_{\max}$ and $\varpi_{\min}$ are the maximum and minimum value of weight parameter, t and T denote current iteration and maximum iteration. In (9), the weight parameter decreases linearly from the maximum value $\varpi_{\max}$ to minimum value $\varpi_{\min}$, which guarantees that all particles hold higher searching speed in the early stage and higher searching accuracy in the later stage.

C. Jointly Optimized Echo State Network

Closely after the L_2 regularization is employed to constraint the output weight W_{out} in (6), the LW-PSO is adopted to optimize the *four* key reservoir parameters (i.e., γ, $\boldsymbol{N}$, *IS*, *SD*). The flaw chat of the proposed JOESN is given in Fig. 1. And the detailed processes are explained as follows.

1) Initialize parameters in LW-PSO: the population size n, accelerate coefficient α_1 and α_2, the maximum iterations T, flight velocity and spatial vector of all particles V and χ, give each particle a random position within a certain range.

2) Initialize parameters and collect data series of L_2-ESN, including penalty coefficients λ, $\boldsymbol{W}_{in}$, $\boldsymbol{W}$ and $\boldsymbol{W}_{back}$, training data $x_{tr}\in\mathbb{R}^{1\times d_{tr}}$ and testing data $x_{te}\in\mathbb{R}^{1\times d_{te}}$, where d_{tr} and d_{te} are the lengths of training data and testing data.

3) Train L_2-ESN according to (1)-(3), (5) and (6), where (5) helps to calculate the fitness value of current particle application in the L_2-ESN model.

4) Iterate for parameter optimization according to (7)-(9), utilize step *3)* to recalculate the fitness value. Compare the current fitness value to the optimal fitness value in history, if it is smaller, update optimal fitness value with current value of (5), if not, then keep optimal fitness value. The *four* optimized parameters are updated under the same manner during the same loop.

5) Judge whether there are iteration ends. If end, jump to step *6)*, if not, jump back to step *4)*.

6) Output the *four* optimal dynamic reservoir parameters for the prediction model.

Herein, we get the final JOESN with optimal parameters, being ready for the upcoming prediction task.

III. Computation Complexity Analysis

Computation complexity is a significant aspect to evaluate the performance for algorithms. As described, the JOESN includes two key aspects, L_2 regularization and LW-PSO. The former is utilized for solving the ill-posed solutions by re-constraining output weights $\boldsymbol{W}_{out}$. The latter is employed to obtain the optimal initial or intermediate parameters in the reservoir of L_2-ESN. According to [7], the computation complexity includes two parts, function operation in (2) and matrix production in (3). For $\boldsymbol{W}_{in}\boldsymbol{U}$, $\boldsymbol{WX}$ and $\boldsymbol{W}_{back}^{T}\boldsymbol{Y}^{T}$, their computation complexities are $O(NML)$, $O(N_{nz}L)$ and $O(NL)$, respectively, where N_{nz} is the number of nonzero elements in $\boldsymbol{W}$. Because $\boldsymbol{W}$ is a sparse matrix, the N_{nz} is generally small. As for the active function in (2), the complexity is $O(NL)$. When it comes to (3), the complexity is $O(L(M+N))$. So the total complexity is $O(\max(NML, N_{nz}L, NL\ L(M+N)))$. As M, N, N_{nz} and L are all positive integers, $NM > N > N_{nz}$, $NM > (N+M)$, so the total complexity of JOESN is $O(NML)$. Because that the output weights $\boldsymbol{W}_{out}$ are learnt in advance but issued only once, while the prediction process is produced relatedly in engineering practice. The total complexity of JOESN can be approximated to that of series prediction, which is at the same level of FESN.

IV. Performance Evaluation of JOESN

A. Evaluation Setup

In this section, *one* open benchmark of 3-D Lorenz chaotic system is used to verify the performance of JOESN. The fundamental ESN (FESN) [6] is taken as the baseline, two state-of-the-art enhanced ESN models (i.e., support vector echo state machine (SVESM) [8] and ridge regression-ESNs (RESN) [9]) are implemented for longitudinal contrastive analysis. Furthermore, two neural-network-based regression algorithms (i.e., Elman network [10] and extreme learning machine (ELM) [11]) are selected for crosswise contrastive analysis. Three landmark criterions are considered for performance evaluation, which are root mean square error (*RMSE*), normalized *RMSE* (*NRMSE*), and symmetric mean absolute percentage error (*SMAPE*) [12].

In order to facilitate readers to reproduce experiments, here we list the related parameters as follows. The tree constants of Lorenz: a=10, b=28, c=8/3; Series: the embedding dimensions are 2, 1, 7 and the time delays are 8, 8, 12 for *x*, *y*, *z* series in Lorenz; Data setup: 4000 samples for training with 100 of them washed out, 1000 samples for test, Data1 are noise-free and Data2 stands for noisy time series suffered with 20 dB Gauss white noise. JOESN: λ=0.0001, n=30, α_1=α_2=1.5, T = 100, $V_{\max}$=[5,1,1,1], $V_{\min}$=[-5,-1,-1,-1], $\chi_{\max}$=[500,1,1,1], $\chi_{\min}$=[1,0.01,0.01,0.01], $\varpi_{\max} = 0.9$, $\varpi_{\min} = 0.2$.

TABLE I. Optimal Parameters in JOESN for Lorenz

Data	γ	***SR***	***N***	***IS***	***SD***
Data1	0.5245	1.1958	454	0.0966	0.0100
Data2	0.9657	1.4468	310	1.0000	0.0100

TABLE II. Prediction Performance of Lorenz

Model	**Data**	***RMSE***	***NRMSE***	***MAPE***
FESN [6]	Data1	0.0618	0.0088	0.0083
SVESM [8]		0.0090	0.0012	0.0054
RESN [9]		0.0100	0.0013	0.0085
ELman [10]		0.1577	0.0205	0.0800
ELM [11]		0.0080	0.0010	0.0060
JOESN		**0.0037**	**0.0005**	**0.0006**
FESN [6]	Data2	1.4657	0.2097	0.0720
SVESM [8]		0.4503	0.0578	0.3485
RESN [9]		0.2114	0.0275	0.1394
ELman [10]		0.6181	0.0802	0.5771
ELM [11]		0.2507	0.0326	0.2239
JOESN		**0.1037**	**0.0148**	**0.0575**

B. Results and Discussions

After the LW-PSO optimization, we obtain the optimal parameters of JOESN listed in Table I. The three evaluation criterions of all the comparative networks for two distinct datasets are listed in Table II. Evidently, whether for noise-free Data1 or noisy Data2, our proposed JOESN shows more competitive accuracy than the other five methods. Notably, the JOESN obtains the best prediction scores with 0.0037 of RMSE, 0.0005 of NRMSE and 0.0006 of MAPE for Data1. When it comes to Data2, all the six competitors experience performance decrease in varying degrees, nevertheless, our proposed JOESN performs also better than all other five competitors. These results prove that JOESN possesses promising anti-noise ability.

V. Experiment Results and Discussions

A. Rayleigh Fading Channel

Considering channel fading of wireless communication as a chaotic process, Sun *et al*. proposed a chaos-based short-term fading channel predictor to obtain CSI [13]. Besides, Jaeger, the inventor of ESN, prophesied that ESN is a promising network for equalizing communication channel [6]. Hence in this subsection, we first apply our proposed JOESN to forecast the Rayleigh fading channel, so as to evaluate its prediction performance. In order to estimate the channel parameter *h(t)* of Rayleigh fading channel, we simulate Rayleigh fading channel by sinusoids superposition method. The *h(t)* is modeled as

$$h(t) = \frac{E_0}{\sqrt{2N_0+1}}\{h_I(t) + jh_Q(t)\} \tag{10}$$

where E_0 is the average amplitude of fading channel. N_0=(N/2-1)/2, N denotes the number of the plane waves. $h_I(t)$ and $h_Q(t)$ denote the fading channel real and imaginary component, respectively, which are estimated by

$$\begin{cases} h_I(t) = 2\sum_{n}^{N_0}(\cos\phi_n \cos w_n t) + \sqrt{2}\cos\phi_N \cos w_d t \\ h_Q(t) = 2\sum_{n}^{N_0}(\sin\phi_n \cos w_n t) + \sqrt{2}\sin\phi_N \cos w_d t \end{cases} \tag{11}$$

TABLE III. OPTIMAL PARAMETERS FOR RAYLEIGH FADING CHANNEL

Data	γ	***SR***	N	***IS***	***SD***
Data1	0.0100	1.0355	478	1.0000	0.4598
Data2	0.0100	0.2417	500	1.0000	0.0100

TABLE IV. OPTIMAL PARAMENTERS FOR RAYLEIGH FADING CHANNEL

Model	**Data**	***RMSE***	***NRMSE***	***MAPE***
AR [14]	Data1	0.0150	0.0365	0.0581
SVM [15]		0.0816	0.1975	0.1532
DWT-AR-LR [16]		0.0175	0.0388	**0.0218**
FESN [6]		0.0216	0.0524	0.0367
JOESN		**0.0127**	**0.0309**	0.0262
AR [14]	Data2	0.1526	0.3577	0.4324
SVM [15]		0.2044	0.4787	0.2930
DWT-AR-LR [16]		**0.0454**	**0.0995**	**0.0761**
FESN [6]		0.1724	0.4069	0.2289
JOESN		0.1334	0.3158	0.1823

where ϕ_n and ϕ_N are the initial phases of n-th Doppler-shifted sinusoid and maximum Doppler frequency f_d, respectively. And the frequency of the Doppler-shifted sinusoid $\{w_n\}_{n=1}^{N_0}$ can be expressed as $w_n = w_d \cos\theta_n \ = 2\pi f_d \cos(2\pi n / N)$, where n=1, 2, 3,…,N_0.

According to the abovementioned method, we simulate several sets of CSI for Rayleigh fading channel with parameters of N_0=8, E_0=1, f_d=500Hz, ϕ_N=0 and sampling period T_s=5*10^{-5}s. And all parameters of JOESN are inherited from Section IV.A. Some classical channel estimation schemes, autoregressive algorithm (AR) [14], support vector machine (SVM) [15] and DWT-based autoregressive and linear regression algorithm (DWT-AR-LR) [16], are selected and implemented for crosswise contrastive analysis. The related parameters are defined as follows. AR: the order is 9; SVM: the type of SVM is epsilon-SVR, the epsilon in loss function of epsilon-SVR is 0.01, and the kernel function type is radial basis function, where the gamma value is 2.8; DWT-AR: the wavelet function is set as bior3.3, the decomposition level is 4, and the orders of LR and AR are 9.

B. Results and Analysis

After learning the optimal parameters of JOESN (Table III), the three evaluation criterions of all the comparative networks for two distinct datasets are listed in Table IV. For Data1, JOESN performs better than its baseline, FESN, and even better than the other three classical methods (i.e., AR, SVM, DWT-AR) in varying degrees, with the nearly best scores of RMSE=0.0127, NRMSE=0.0309 and SMAPE=0.0262. When it comes to Data2, after addressing the noise sensitive problem of FESN by importing L_2 regularization, the JOESN exhibits better noise-robustness (Fig. 2), producing higher scores than classical AR and SVM. However, DWT-AR-LR holds the best performance. The in-depth reason is that Data2 is decomposed into various components by discrete wavelet transform (DWT). And DWT is a noise-robust transformation, it can support DWT-AR-LR working better by avoiding noisy components selectively. Nevertheless, our JOESN is expected to catch up with or even exceed DWT-AR-LR if more appropriate embedding dimension and time delay could be set. Consequently, it is urgent to build a methodology to adaptively adjust these two vital settings in signal reconstruction stage by investigating the influencing mechanisms from temporal complexity and signal quality of time series in future work.

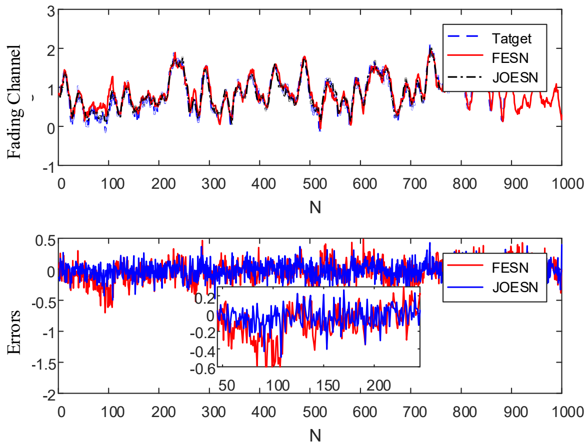

Fig. 2. Prediction performance of FESN and JOESN for noisy Data2.

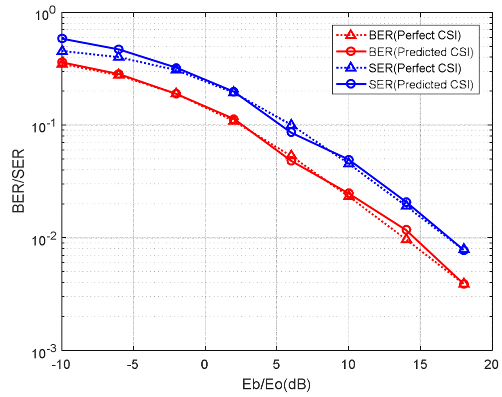

Fig. 3. BER/SER evaluation when using JOESN for CSI prediction.

For further understanding, we continue to evaluate the data transmission quality of the wireless channel by estimating its bit error rate (BER) and symbol error rate (SER). For general purpose, 2-ary differential phase-shift keying (2-DPSK) is employed for signal modulation at the transmitter, and coherent demodulation is introduced for signal detection at the receiver. And to avoid any setup modification, the JOESN is directly selected from Table IV to predict a Rayleigh fading channel. Comparative testing results of BER and SER for the perfect CSI and predicted CSI are given in Fig.3. The lateral axis, Eb/N0 (dB) denotes the power spectral density ratio of transmitted bit streams to channel noises. It is fairly clear that the main trend of these four curves decease with the increase of Eb/N0, among of them, the BER curve with CSI prediction assisted by JOESN is very close to the theoretical BER curve. When it comes to the SER, except for a slight performance degradation when Eb/N0 is extremely low (<-10dB), in most cases, the communication quality under our prediction scheme is close to that of the ideal channel with the expected CSI.

VI. CONCLUSION

In order to improve the prediction accuracy and noise robustness, a jointly optimized echo state network (JOESN) is proposed based on output weight regularization and initial parameter optimization. In this prediction model, ill-posed solutions can be avoided to a large extent due to the sufficiently unbiased and small-amplitude output weights regularized by L_2 penalty in the loss function. The employed linear-weighted particle swarm optimization (LW-PSO) permits the JOESN intelligently learn its reservoir parameters. This automatic strategy for parameter optimization prevent our learning networks from inefficiency and unreliability of manual settings. The simulation tests on Lorenz chaotic system prove that our proposed JOESN performs better especially in the case of noises. Furthermore, the proposed JOESN yields acceptable precision and reliability in a preliminary application using a Rayleigh fading channel dataset. These positive results indicate that our proposed JOESN has considerable potential in obtaining CSI of fading channel, which shows a promising application prospect in wireless communications.

ACKNOWLEDGMENT

This research was jointly supported in part by the National Natural Science Foundation of China under Grant 61973323 and Grant 51704089, and the Anhui Provincial Natural Science Foundation of China under Grant 1808085QF190.

REFERENCES

[1] M. Mohammadkarimi, E. Karami, O. A. Dobre, and M. Z. Win, “Doppler spread estimation in MIMO frequency-selective fading channels,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1951–1965, Mar. 2018.

[2] L. Cheng, Y.-C. Wu, J. Zhang, and L. Liu, “Subspace identification for DOA estimation in massive/full-dimension MIMO systems: Bad data mitigation and automatic source enumeration,” *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5897–5909, Nov. 2015.

[3] R. Shafin, L. Liu, J. Ashdowny, J. Matyjasy, M. Medleyy, B. Wysockiy, and Y. Yi, “Realizing green symbol detection via reservoir computing: An energy-efficiency perspective,” in *IEEE ICC 2018*, Kansas City, MO, USA, May 2018, pp. 1-6.

[4] K. Burse, R. N. Yadav, and S. C. Shrivastava, “Channel Equalization Using Neural Networks: A Review,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 40, no. 3, pp. 352–357, May 2010.

[5] S. Mosleh, L. Liu, C. Sahin, Y. Zheng and Y. Yi. “Brain-inspired wireless communications: Where reservoir computing meets MIMO-OFDM,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 1-15, Dec. 2017.

[6] H. Jaeger and H. Haas, “Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication,” *Science*., vol. 304, no. 5667, pp. 78-80, Apr. 2004.

[7] Y. Zhao, H. Gao, N. C. Beaulieu, Z Chen.and H Ji, “Echo state network for fast channel prediction in Ricean fading scenarios,” *IEEE Commun. Lett.*, vol. 21, pp. 672-675, Mar. 2017.

[8] Z. W. Shi and M Han, “Support vector echo-state machine for chaotic time-series prediction.” IEEE Trans. Neural Netw., vol. 18, no. 2, pp. 359-372, Mar. 2007.

[9] X. Dutoit, B. Schrauwen, J. V. Campenhout, et al. “Pruning and regularization in reservoir computing,” Neurocomputing., vol. 72, no. 7, pp. 1534–1546, Mar. 2009.

[10] M. C. Ozturk, D. Xu and J. C. Príncipe, “Analysis and design of echo state networks,” *Neural Comput*., vol. 19, no. 1, pp. 111-138, Jan. 2007.

[11] M. Li, G. Huang, P. Saratchandran and N. Sundararajan, “Fully complex extreme learning machine,” *Neurocomputing*., vol. 68, no. 1, pp. 306-314, Oct. 2005.

[12] C. bal and S. demir, “Adaptive weighted performance criterion for artificial neural networks,” in *IEEE IDAP 2018*., pp. 28-30, Sep. 2018.

[13] Z. Xiang, T. Zhang and J. Sun, “Prediction algorithm for fast fading channels based on the chaotic attractor,” *J. Xidian Univ.*, vol. 33, no. 1, pp. 145-149, Jan. 2006.

[14] Y. Sui, W. Yu and Q. Luo, “Jointly optimized extreme learning machine for short-term prediction of fading channel,” *IEEE Access*, vol. 6, pp. 1-11, 2018.

[15] N. Zhao, C. Hou and Q. Wang, “A new SVM-based modeling method of cabin path loss prediction,” *Int. J. Antennas Propag.*, vol. 2013, no. 2, pp. 1-7, May. 2013.

[16] X. Long and B. Sikdar, “A Wavelet Based Long Range Signal Strength Prediction in Wireless Networks,” in *IEEE Intl. Conf. Commun.*, Beijing, China, 2008, 2043-2047.