IOP Conf. Series: Materials Science and Engineering **435** (2018) 012025 doi:10.1088/1757-899X/435/1/012025

Water Quality Prediction in the Waste Water Treatment Process Based on Ridge Regression Echo State Network

Jing Zhao[1,2,*], Chao Zhao[1,2], Fan Zhang[1,2], Gang Wu[1,2] and Haitao Wang[1,2]

[1] China National Institute of Standardization, Beijing 100191, China
[2] AQSIQ Key Laboratory of Human Factors and Ergonomics(CNIS), Beijing 100191, China

[*]E-mail: nlzhaojing0506@163.com

Abstract. Echo state network (ESN), a novel recurrent neural network, has a randomly and sparsely connected reservoir. Since the output weights are computed by Moore-Penrose inverse, the ill-posed problem may exist in the ESN. To overcome this problem, ridge regression echo state network (RESN) is proposed, in which the ridge regression algorithm is used to calculate the output weights instead of linear regression. Simulation results show that the RESN has better performance than some other existing methods, thus can deal with the ill-posed problem.

1. Introduction

Water quality prediction in the wastewater treatment process (WWTP) can provide strong support for water treatment plant management decisions [1-2]. However, it is very difficult to predict the water quality, since WWTP is a complex system including a variety of physical and biochemical reactions. Moreover, due to the nonlinear characteristics, delay-time and uncertainty, it is difficult to measure effluent qualities parameters in the WWTP. The quality of treated wastewater is measured by some parameters, such as biochemical oxygen demand (BOD), total phosphorous (TP), ammonia nitrogen (NH_4-N), and so on, which must meet the national standard. The traditional measurement approaches usually depend on the laboratory analysis which takes a long time [3-4]. For example, it takes about five days to get BOD values according to the conventional chemical measurement approaches, which cannot have a real time monitoring process of the water pollution situation. This lack of real- time process variable information limits the effective operation of effluent water quality prediction. Furthermore, the online monitoring instrument needs high economic costs and is difficult to be conducted in the WWTP. Therefore, water quality prediction model is essential to support water quality parameters.

To solve the above-mentioned problems, soft computing method has been widely used in complex system modelling, where the hard-to-measure process variables are estimated by the other easy-to-measure variables. The artificial neural networks (ANNs) are significant nonlinear approaches for nonlinear system modelling and have attracted a lot of attention. In [5], Takens embedding theorem based phase space reconstruction is used to extract more information from the limited datasets of the chaotic system, where principal component analysis and artificial neural network is then adopted to estimate the effluent BOD value. In [6], an improved T–S fuzzy neural network (TSFNN) is used to predict BOD values based on the knowledge representation ability and learning capability, where a gradient descent method with the momentum item is used to adjust antecedent parameters and consequent parameters. In [2], a flexible structure radial basis function neural network (FS-RBFNN) is proposed for water quality prediction, where the hidden neurons in the RBF neural network can be added or removed online based on the neuron activity and mutual information. In [1], a self-organizing cascade neural network (SCNN) with random weights is proposed for water quality prediction on effluent BOD and TP. The SCNN is constructed via simultaneous structure and parameter learning processes. The simulation results show that the proposed SCNN has better performance.

Recurrent neural network (RNN) is a significant nonlinear approach for nonlinear system modelling and has attracted increasingly more and more attention. However, the traditional RNN based on the gradient method has gradient disappearance and gradient explosion problem. As an effective alternative for RNN training, echo state network (ESN) [7] can overcome the local minima and gradient vanishing problems. A typical ESN has a large and sparsely connected reservoir. During the training process, only the output weights are computed by simple Moore-Penrose inverse method, while the input weights and reservoir weights remain fixed once they are generated. This simple and effective training approach makes ESN have numerous successful applications such as time-series prediction [8], speech recognition [9], and nonlinear signal processing [10].

Although ESN has better performance than the traditional neural network, there are still some problems in the ESN. Since the output weights are trained by the simple regression, the ill-posed problem may exist in the training process. If it happens, the large output weights might occur and the generalization ability will be degraded. Additionally, when the number of training samples is less than the number of reservoir neurons, the ill-posed problem must happen. To solve the ill-posed problem, noise injection is used to improve the network performance, however, there are no satisfactory methods to determine the amplitude of noise. The intrinsic plasticity rule is used in the reservoir computing setting [11]. However, parameter adaption process based on the gradient method is complicated and some hyper-parameters are obtained by experience. In this paper, ridge regression method is used to address the ill-posed problem, where the regularization term is introduced into the loss function to penalize larger readout weights and the ridge parameter is determined by the cross-validation method.

The remainder of this paper is organized as follows: In section 2, a brief review of the original ESN model is given. In section 3, the ridge regression echo state network model is proposed. In section 4, the stability analysis is given. In section 5, the simulations are carried out to illustrate the performance of the proposed algorithm. Finally, some conclusions are drawn.

2. Echo state network

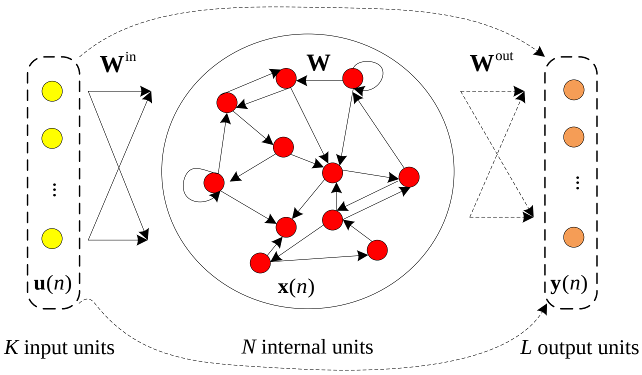

Figure 1. The basic architecture of OESN without feedback connections.

An original ESN (OESN) [7] without feedback connections consists of three layers, input layer, reservoir layer and readout layer. The structure of OESN is presented in figure 1. It is assumed that the OESN has K input neurons, N reservoir neurons and L readout neurons, respectively. $\mathbf{u}(n)$, $\mathbf{x}(n)$, $\mathbf{y}(n)$ denote the input vector, the reservoir state vector and the readout vector, respectively. The updating formula of OESN is given as follows:

$$\mathbf{x}(n) = \mathbf{tanh}(\mathbf{W}^{in}\mathbf{u}(n) + \mathbf{W}\mathbf{x}(n-1)) \;, \tag{1}$$

$$\mathbf{y}(n) = \mathbf{W}^{out}(\mathbf{u}(n), \mathbf{x}(n)) \quad , \tag{2}$$

Where $\mathbf{W}^{in}$, $\mathbf{W}$, $\mathbf{W}^{out}$ are the input weight matrix, reservoir weight matrix and output weight matrix, respectively. $\mathbf{f}$ is the reservoir active function, which is usually chosen as hyperbolic tangent function. The output weights are computed by using the linear regression method as follows

$$\mathbf{W}^{out} = ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z})^T \tag{3}$$

Where $\mathbf{X}=[\mathbf{X}(1),\mathbf{X}(2),\cdots,\mathbf{X}(P)]^T$ (P is the number of training samples) denotes the internal state matrix, $\mathbf{X}(n)= [\mathbf{u}(n)^T, \mathbf{x}(n)^T]^T$, $\mathbf{Z}=[\mathbf{z}(1),\mathbf{z}(2),\cdots,\mathbf{z}(P)]^T$ is the desired output.

3. Ridge regression echo state network

From formula (3), if $\mathbf{X}^T\mathbf{X}$ is singular, the ill-posed problem will happen. To obtain a stabilized solution, the ridge regression, a widely used regularization method, is applied to introduce a penalty term into the loss function as follows.

$$J(\mathbf{W}^{out}) = \arg\min_{\mathbf{W}^{out}} \left\|\mathbf{Z} - \mathbf{X}\mathbf{W}^{out}\right\|_2^2 + \lambda\left\|\mathbf{W}^{out}\right\|_2^2 \tag{4}$$

Where $\lambda > 0$ denotes the ridge parameter. Differentiating with respect to $\mathbf{W}^{out}$, it can be got as follows

$$\frac{\partial J(\mathbf{W}^{out})}{\partial \mathbf{W}^{out}} = -2\mathbf{X}^T(\mathbf{Z} - \mathbf{X}\mathbf{W}^{out}) + 2\lambda\mathbf{W}^{out} \tag{5}$$

Let $\frac{\partial J(\mathbf{W}^{out})}{\partial \mathbf{W}^{out}} = 0$, we have $\mathbf{X}^T\mathbf{X}\mathbf{W}^{out} - \mathbf{X}^T\mathbf{Z} + \lambda\mathbf{W}^{out} = 0$, the following equation can be obtained

$$(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})\mathbf{W}^{out} = \mathbf{X}^T\mathbf{Z} \tag{6}$$

Since the matrix $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}$ must be invertible, the solution of (4) is obtained as follows

$$\mathbf{W}^{out} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{Z} \;. \tag{7}$$

Where ridge parameter is obtained by cross-validation method. From (7), it is known that the ill-posed problem can be solved.

The detailed steps are summarized in Algorithm 1.

Algorithm 1

Step 1: Randomly generate an input weight matrix $\mathbf{W}^{in}$ according to uniform distribution and initialize the reservoir states $\mathbf{X}$ (0).

Step 2: Randomly generate a reservoir weight matrix $\mathbf{W}_0$ with the predefined sparsity and reservoir size. Scale $\mathbf{W}_0$ to $\mathbf{W}= (\alpha_w/ \rho(\mathbf{W}_0))\ \mathbf{W}_0$, where $0<\alpha_w<1$ and $\rho(\mathbf{W}_0)$ is the spectral radius of $\mathbf{W}_0$.

Step 3: Drive the reservoir by the input signals as (1), discarded a certain number of initial steps and collect the internal states at an initial transient $n_{\min}$.

Step 4: Compute the output weight matrix $\mathbf{W}^{out}$ using (7).

Step 5: Test the trained RESN.

4. Stability analysis

The key of the ESN is that the reservoir should have the echo state property (ESP). ESP means that the internal states should be uniquely depend on external input and it is related to the input samples and reservoir weight. To describe the ESP, the local dynamics of the system by linearizing the RESN is considered. The nonlinear system (1) can be approximated as follows

$$\mathbf{x}(n) = \mathbf{g}'\,\mathbf{W}\mathbf{x}(n-1) + \mathbf{g}'\,\mathbf{W}^{in}\mathbf{u}(n) \triangleq \mathbf{A}\mathbf{x}(n-1) + \mathbf{B}\mathbf{u}(n) \tag{8}$$

where $\mathbf{g}' = \mathbf{tanh}', \left\|\mathbf{g}'\right\| \le 1$, denote $\mathbf{A} = \mathbf{g}'\,\mathbf{W}, \mathbf{B} = \mathbf{g}'\,\mathbf{W}^{in}$.

The sufficient condition of the ESP is that the maximal singular value of reservoir weight matrix $\mathbf{W}$ is less than 1 ($\sigma(\mathbf{W}) < 1$). Since $\left\|\mathbf{W}\right\| = \sigma(\mathbf{W})$, the sufficient condition is equivalent to $c \triangleq \left\|\mathbf{W}\right\| < 1$.

Suppose $\mathbf{x}(n)$ and $\mathbf{x}'(n)$ are different internal state vectors.

$$\begin{aligned}\|\mathbf{x}(n)-\mathbf{x}'(n)\| &= \|\mathbf{Ax}(n-1)+\mathbf{Bu}(n)-\mathbf{Ax}'(n-1)-\mathbf{Bu}(n)\| = \|\mathbf{Ax}(n-1)-\mathbf{Ax}'(n-1)\| \\ &\le \|\mathbf{A}\|\|\mathbf{x}(n-1)-\mathbf{x}'(n-1)\| \le \|\mathbf{g}'\|\cdot\|\mathbf{W}\|\cdot\|\mathbf{x}(n-1)-\mathbf{x}'(n-1)\| \le c\|\mathbf{x}(n-1)-\mathbf{x}'(n-1)\| \\ &\le \cdots \le c^n\|\mathbf{x}(0)-\mathbf{x}'(0)\|.\end{aligned} \tag{9}$$

This shows that the current reservoir state is determined by its past external input history, which guarantees the ESP.

5. Simulation

In this section, the performance of the RESN is evaluated on Lorenz time series prediction, effluent BOD and TP prediction in the WWTP. To show the effectiveness of the RESN, the simulations are compared with the following models: OESN [7], TSFNN [6], and FS-RBFNN [2]. The reservoir size, spectral radius, and sparsity are selected as 200, 0.85 and 0.05, respectively. All simulations are tested in MATLAB 2013b environment and run on i7-4790 with CPU 3.60GHz and 8.0GB RAM.

The normalized root mean square error (NRMSE) is used as the evaluation criteria of model performance, which is defined as below:

$$\text{NRMSE} = \sqrt{\sum_{t=1}^{S}\frac{(d_i(t)-y_i(t))^2}{S\sigma^2}} \tag{10}$$

Where $d_i(t)$ denotes the desired output, $y_i(t)$ is the corresponding prediction output, σ^2 is the variance of the desired outputs, and S is the total number of $d_i(t)$.

5.1. Lorenz time series prediction

The Lorenz system is a mathematical model for atmospheric convection, which is widely used as a benchmark in many applications [12]. It can be described as follows

$$\begin{cases}\dot{x} = a_1(y-x) \\ \dot{y} = -xz + a_2x - y. \\ \dot{z} = xy - a_3z\end{cases} \tag{11}$$

Where a_1=10, a_2=28 and a_3=8/3 are the system parameter. Figure 2 gives the the trajectory of the Lorenz system.

In this experiment, the data set are generated by the fourth-order Runge-Kutta method with a step size 0.01, and only the *Y*-dimension samples y(t) are used for time series prediction. For y(t), 5000 data samples are generated, the first 3000 values are used for training, and the last 2000 values are used to test the proposed model. In the training set, the first 1000 samples are discarded to washout the initial transient. The initial value is selected as $x(0)=1$, $y(0)=1$, $z(0)=0$.

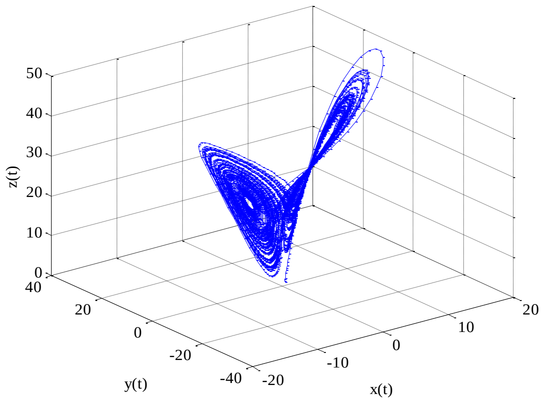

Figure 2. Lorenz attractor

Figure 3 and figure 4 give the testing output and testing error, respectively. Obviously, the proposed

RESN has better prediction performance than OESN. The testing error of RESN is limited in [-0.3×10^{-4}, 0.3×10^{-4}], while the testing error of OESN is limited in [-0.9×10^{-4}, 0.9×10^{-4}].

The detailed comparison results are listed in Table 1 based on the average of 50 independent simulations. As listed in Table 1, although RESN takes more training time than OESN, it has less training NRMSE and testing NRMSE compared with other models.

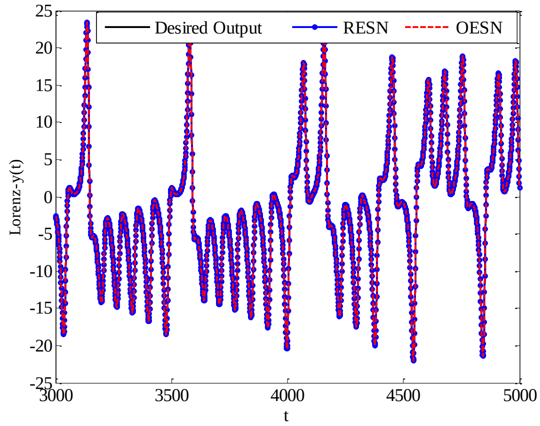

Figure 3. Testing output for Lorenz time series

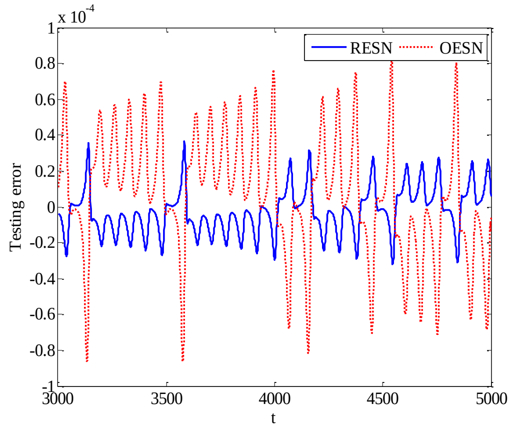

Figure 4. Testing error for Lorenz time series

Table 1. Comparison of different models for Lorenz time series

Models	Training time (s)	Training NRMSE	Testing NRMSE
OESN[7]	**35.18**	4.18×10^{-5}	8.36×10^{-4}
TSFNN[6]	48.65	6.21×10^{-4}	7.68×10^{-3}
FS-RBFNN[2]	65.32	8.15×10^{-4}	8.96×10^{-3}
RESN	41.18	$\mathbf{3.28\times10^{-5}}$	$\mathbf{5.28\times10^{-4}}$

5.2. Water quality prediction in the WWTP

In recent years, wastewater has become one of the major environment problems in the national governments worldwide. Treating wastewater and predicting water quality have become very important [1]. However, the municipal sewage treatment system is a typical nonlinear system, which is difficult to be modeled for complex biochemical reaction. In this section, the main purpose is to provide accurate predictions of effluent BOD and TP, which are one of the most important effluent quality indices and can reflect the water pollution situation.

Table 2. Variables used in the prediction model

Output	Inputs
Effluent BOD	COD; TSS; pH; DO
Effluent TP	Influent TP; T; ORP; DO; TSS; pH

Table 3. List of acronyms used in simulations

Acronyms	Description
BOD	Biochemical oxygen demand
COD	Chemical oxygen demand
TP	Total phosphorus
TSS	Total suspended solids
DO	Dissolved oxygen
ORP	Oxidation reduction potential
T	Temperature

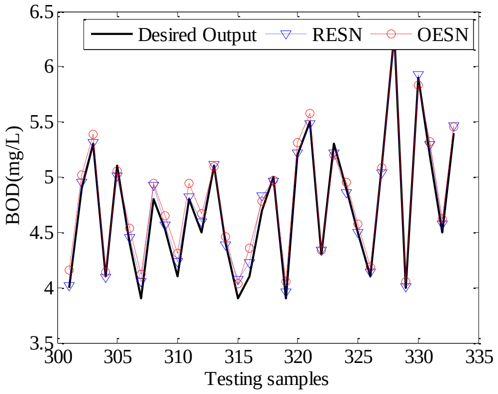

Figure 5. Testing outputs for effluent BOD

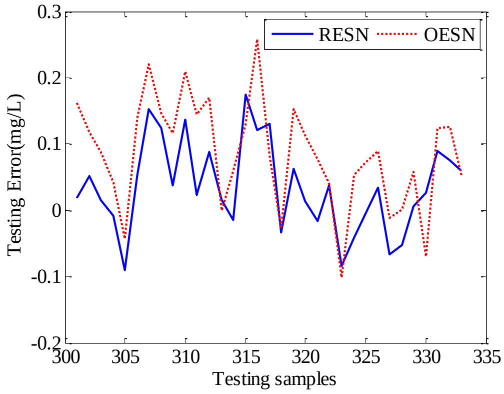

Figure 6. Testing error for effluent BOD

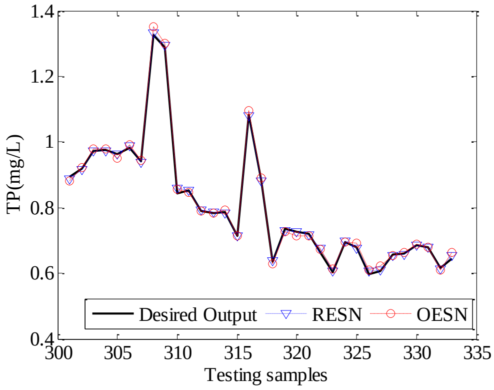

Figure 7. Testing outputs for effluent TP

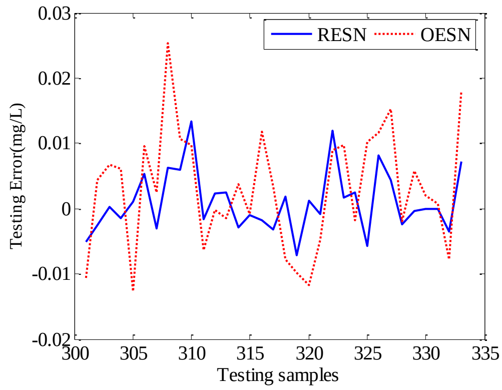

Figure 8. Testing error for effluent TP

Table 4. Comparison of different models for effluent BOD prediction

Models	Training time (s)	Training NRMSE	Testing NRMSE
OESN[7]	**25.13**	0.1742	0.1956
TSFNN[6]	79.65	0.3328	0.4376
FS-RBFNN[2]	93.32	0.3129	0.4267
RESN	32.18	**0.1182**	**0.1329**

Table 5. Comparison of different models for effluent TP prediction

Models	Training time (s)	Training NRMSE	Testing NRMSE
OESN[7]	**24.36**	0.0032	0.0156
TSFNN[6]	78.63	0.0098	0.0398
FS-RBFNN[2]	88.35	0.0069	0.0278
RESN	31.19	**0.0018**	**0.0089**

After deleting the abnormal data, 333 samples are obtained from a sewage treatment plant in Beijing, China. The RESN selects some important subsidiary variables for effluent BOD and TP prediction, respectively. The detailed water quality parameters listed in Table 2 are selected as the input variables of the prediction model, and the description of each variable can be found in Table 3.

For the samples data, the first 200 samples are used for training, 50 samples in training set are discarded to washout initial transient, and the last 133 values are used to test the network performance. Before the simulation, the inputs and the target outputs are normalized into [-1, 1]. After simulations, the outputs are converted.

The simulation results for effluent BOD are shown in figure 5 and figure 6, respectively, which illustrate that the RESN has more accurate prediction than OESN for actual time-series. Likewise, the simulation results for effluent TP are shown in figure 7 and figure 8, respectively. Based on 50 independent simulations, the detailed results for effluent BOD and TP are listed in Table 4 and Table 5, respectively. From the comparison results of different models, it is seen that the training time of RESN is more than OESN, however, RESN is faster than TSFNN and FS-RBFNN. The training NRMSE and the testing NRMSE of RESN are less than other models, which means that the RESN has high accuracy than other models.

6. Conclusion

In this paper, an RESN model based on ridge regression echo state network is proposed. Since the output weights are computed by Moore-Penrose inverse, the ill-posed problem may exist in the ESN. The proposed RESN can solve the ill-posed problem by adding a penalty term to the loss function. The simulation results show that the proposed RESN model can get good performance and has a smaller testing error than other models.

Acknowledgments

This research was supported by China National Institute of Standardization (522018Y-5941, 522018Y-5948).

References

[1] Li F J, Qiao J F, Han H G, Yang C L 2016 A self-organizing cascade neural network with random weights for nonlinear system modeling *Appl. Soft Comput.* **42** 184
[2] Han H G, Chen Q L, Qiao J F 2011 An efficient self-organizing RBF neural network for water quality prediction *Neural Netw.* **24**(7) 717
[3] Han H G, Qiao J F 2010 A self-organizing fuzzy neural network based on a growing-and-pruning algorithm *IEEE Trans. Fuzzy Syst.* **18**(6) 1129
[4] Han H G, Qiao J F 2013 A structure optimisation algorithm for feedforward neural network construction *Neurocomputing* **99** 347
[5] Qiao J F, Hu Z Q, Li W J 2016 Soft measurement modeling based on chaos theory for biochemical oxygen demand (BOD) *Water* **8** 581
[6] Qiao J F, Li W, Han H G 2014 Soft computing of biochemical oxygen demand using an improved T–S fuzzy neural network *Chin. J. Chem. Engin.* **22**(11-12) 1254
[7] Jaeger H, Haas H 2004 Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication *Science* **304**(5667) 78
[8] Qiao J F, Li F J, Han H G, Li W J 2017 Growing echo-state network with multiple subreservoirs *IEEE Trans. Neural Netw. Learn. Syst.* **28**(2) 391
[9] Skowronski M D, Harris J G 2007 Automatic speech recognition using a predictive echo state network classifier *Neural Netw.* **20**(3) 414
[10] Xia Y, Jelfs B, Hulle M M V 2011 An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals *IEEE Trans. Neural Netw.* **22**(1) 74
[11] Schrauwen B, Wardermann M, Verstraeten D 2008 Improving reservoirs using intrinsic plasticity *Neurocomputing* **71**(7–9) 1159
[12] Xu M, Han M 2016 Adaptive elastic echo state network for multivariate time series prediction *IEEE Trans. Cybern.* **46**(10) 2173