An Augmented Echo State Network for Nonlinear Adaptive Filtering of Complex Noncircular Signals

Yili Xia, *Student Member, IEEE*, Beth Jelfs, Marc M. Van Hulle, *Senior Member, IEEE*, José C. Príncipe, *Fellow, IEEE*, and Danilo P. Mandic, *Senior Member, IEEE*

***Abstract*— A novel complex echo state network (ESN), utilizing full second-order statistical information in the complex domain, is introduced. This is achieved through the use of the so-called augmented complex statistics, thus making complex ESNs suitable for processing the generality of complex-valued signals, both second-order circular (proper) and noncircular (improper). Next, in order to deal with nonstationary processes with large nonlinear dynamics, a nonlinear readout layer is introduced and is further equipped with an adaptive amplitude of the nonlinearity. This combination of augmented complex statistics and enhanced adaptivity within ESNs also facilitates the processing of bivariate signals with strong component correlations. Simulations in the prediction setting on both circular and noncircular synthetic benchmark processes and real-world noncircular and nonstationary wind signals support the analysis.**

***Index Terms*— Augmented complex statistics, complex noncircularity, echo state networks, widely linear modeling, wind prediction.**

I. Introduction

RECURRENT neural networks (RNNs) are a class of nonlinear adaptive filters with feedback, whose computational power stems from their ability to act as universal approximators for any continuous function on a compact domain [1], [2]. Owing to their rich inherent memory through feedback, RNNs have found applications in the modeling of highly nonlinear dynamic systems and the associated attractor dynamics. They are typically used in the system identification [3], [4], time-series prediction [5], [6], and adaptive noise cancellation settings [7], [8], where for the nonstationary and nonlinear nature of the signals and typically long impulse responses, using the class of static feedforward networks or transversal filters would result in undermodeling [2], [9], [10].

Recently, a class of discrete-time RNNs, called echo state networks (ESNs), have been introduced, with the aim to reduce the complexity of computation encountered by standard RNNs [11]. The principle behind ESNs is to separate the RNN architecture into two constituent components: a recurrent architecture, called the "dynamical reservoir" or "hidden layer," and a memoryless output layer, called the "readout neuron." The recurrent architecture consists of a randomly generated group of hidden neurons with a specified degree of recurrent connections, and should satisfy the so-called "echo state property" to maintain stability [12]. This way, the high computational complexity of RNNs is significantly reduced due to the sparse connections among the hidden neurons, in addition, the learning requirements are reduced to only the weights connecting the hidden layer and the readout neuron.[1]

Many real-world bivariate processes, such as vector fields and directional signals with "intensity" and "direction" components, are most conveniently represented when considered complex-valued [13]. Consequently, in the neural network literature, several important approaches have been extended to the complex domain, examples include coherent neural networks for sensorimotor systems [14], sonar signal prediction and image enhancement by multivalued neurons [15], gray-scale image processing by complex-valued multistate neural associate memory [16], and geometric figure transformation via complex-valued backpropagation networks [17].

The first extension of ESNs into the complex domain $\mathbb{C}$ was proposed in [18], this network had a linear output mapping, and was trained by the complex-valued Wiener filter, thus making the network second-order optimal for the processing of circular stationary data. Results in adaptive filtering dealing with real-world complex-valued data suggest that, due to the linearity of the output mapping, the degree of universal function approximation exhibited by standard ESNs may not be sufficient. To that end, a nonlinear output layer within ESNs, i.e., the linear mapping followed by a nonlinear activation function, has been proposed in [19]. To deal with common problems experienced in neural network training, such as saturation and slow convergence resulting from the unknown and large dynamics of inputs, the nonlinear output layer of ESNs has further been equipped with an adaptive amplitude of the nonlinearity [19].

Adaptive filtering algorithms in the complex domain $\mathbb{C}$ are usually considered generic extensions of their real domain counterparts. For instance, a common assumption explicitly

Manuscript received November 12, 2009; revised August 1, 2010 and September 29, 2010; accepted October 2, 2010. Date of publication November 11, 2010; date of current version January 4, 2011.

Y. Xia, B. Jelfs, and D. P. Mandic are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: yili.xia06; beth.jelfs05; d.mandic@imperial.ac.uk).

M. M. Van Hulle is with the Laboratorium Neuro- en Psychofysiologie, Katholieke Universiteit Leuven, Leuven 3000, Belgium (e-mail: marc@neuro.kuleuven.be).

J. C. Príncipe is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: principe@cnel.ufl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNN.2010.2085444

[1] A recent special issue of *Neural Networks*, vol. 20, no. 3, 2007, edited by H. Jaeger, W. Maass, and J. C. Príncipe, was dedicated solely to ESNs and liquid state machines.

or implicitly exists in the signal processing literature that complete second-order statistical information of a zero-mean complex vector $\mathbf{z}$ is contained in the covariance matrix $E[\mathbf{z}\mathbf{z}^H]$. However, recent results in so-called augmented complex statistics show that, in general, this leads to suboptimal estimation [20] and that for the generality of complex-valued random processes both the covariance matrix $E[\mathbf{z}\mathbf{z}^H]$ and the pseudo-covariance matrix $E[\mathbf{z}\mathbf{z}^T]$ should be considered to completely capture the second-order statistics available. In practice, this is achieved by widely linear modeling [21], which has been proved to be particularly advantageous when processing second-order noncircular (improper) signals for which the probability distributions are not rotation invariant[2] [13], [22].

Recently, augmented complex statistics have been introduced into several key learning algorithms, examples include the augmented complex least means square (ACLMS) [23], augmented complex extended Kalman filter [24], and augmented complex real-time recurrent learning [25]. Following on these results, we here introduce augmented statistics into the training of complex ESNs, allowing us to make use of all the available second-order statistical information, and to produce optimal estimates for second-order noncircular (improper) signals.

This paper is organized as follows. In Section II, we provide an overview of widely linear estimation and second-order augmented complex statistics. In Section III, the augmented complex ESN and its nonlinear variants are derived. Simulations on both synthetic circular and noncircular signals and real-world nonstationary and noncircular wind signals are given in Section IV, demonstrating the advantage of the augmented ESN over standard complex ESNs. Finally, Section V concludes this paper.

II. Widely Linear Modeling

Consider the real-valued mean squared error (MSE) estimator

$$\hat{y} = E[y|x] \tag{1}$$

which estimates the values of signal y in terms of another observation x. For zero-mean jointly normal y and x, the linear model solution is

$$\hat{y} = \mathbf{x}^T\mathbf{h} \tag{2}$$

where $\mathbf{h} = [h_1, \ldots, h_N]^T$ is a vector of fixed filter coefficients, and the past of the observed variable is contained in the regressor vector $\mathbf{x} = [x_1, \ldots, x_N]^T$.

In the complex domain, it is assumed that we can use the same form of conditional mean estimator that for real-valued signals given in (1), leading to the standard complex linear minimum mean squared error (MMSE) estimator[3]

$$\hat{y} = \mathbf{z}^H\mathbf{h} \tag{3}$$

where the symbol $(\cdot)^H$ denotes the Hermitian transform operator. However, the real-valued linear estimator in (2) applies to both the real and imaginary parts of complex variables

$$\begin{aligned} \hat{y}_r &= E[y_r|z_r, z_i] \\ \hat{y}_i &= E[y_i|z_r, z_i]. \end{aligned} \tag{4}$$

A more general MSE estimator than that in (3) can be expressed as

$$\hat{y} = E[y_r|z_r, z_i] + \jmath E[y_i|z_r, z_i]. \tag{5}$$

Upon employing the identities $z_r = (z + z^*)/2$ and $z_i = (z - z^*)/2\jmath$ we arrive at

$$\hat{y} = E[y_r|z, z^*] + \jmath E[y_i|z, z^*] \tag{6}$$

leading to a *widely linear* estimator for complex-valued data, given by

$$y = \mathbf{z}^T\mathbf{h} + \mathbf{z}^H\mathbf{g} \tag{7}$$

where $\mathbf{h}$ and $\mathbf{g}$ are complex-valued coefficient vectors. This estimator is suitable for linear MMSE estimation of the generality of complex-valued processes (both circular and noncircular) [21], as it accounts for complete second-order information in $\mathbb{C}$, as shown below.

A. Second-Order Augmented Complex Statistics

From (7), it is clear that the covariance matrix $\mathrm{C}_{\mathbf{zz}} = E[\mathbf{z}\mathbf{z}^H]$ alone does not have sufficient degrees of freedom to describe full second-order statistics in $\mathbb{C}$ [20] and, in order to make use of all the available statistical information, we also need to consider the pseudo-covariance matrix $\mathrm{P}_{\mathbf{zz}} = E[\mathbf{z}\mathbf{z}^T]$. Processes whose second-order statistics can be accurately described by only the covariance matrix, i.e., those for which the pseudo-covariance $\mathrm{P}_{\mathbf{zz}} = \mathbf{0}$, are termed second-order circular (or proper). In general, the notion of circularity extends beyond second-order statistics to describe the class of signals with rotation-invariant distributions $\mathcal{P}[\cdot]$ for which $\mathcal{P}[\mathbf{z}] = \mathcal{P}[\mathbf{z}e^{j\theta}]$ for $\theta \in [0, 2\pi)$. In most real-world applications, complex signals are second-order noncircular or improper, and their probability density functions are not rotation-invariant. In practice, to account for the improperness, the input vector $\mathbf{z}$ is concatenated with its conjugate $\mathbf{z}^*$, to produce an augmented $2N \times 1$ input vector

$$\mathbf{z}^a = \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix}. \tag{8}$$

This augmented input, together with the augmented weights $\mathbf{w}^a = [\mathbf{h}^T, \mathbf{g}^T]^T$, forms a widely linear estimate in (7), and its $2N \times 2N$ augmented covariance matrix is given by [22]

$$\mathrm{C}_{\mathbf{z}^a\mathbf{z}^a} = E\begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix}\begin{bmatrix} \mathbf{z}^H\mathbf{z}^T \end{bmatrix} = \begin{bmatrix} \mathrm{C}_{\mathbf{zz}} & \mathrm{P}_{\mathbf{zz}} \\ \mathrm{P}^*_{\mathbf{zz}} & \mathrm{C}^*_{\mathbf{zz}} \end{bmatrix}. \tag{9}$$

This matrix now contains the complete complex second-order statistical information available in the complex domain, (see [13], [27] for more details).

[2]Circular complex processes have rotation-invariant probability distribution functions.

[3]Both $y = \mathbf{z}^T\mathbf{h}$ and $y = \mathbf{z}^H\mathbf{h}$ are correct, yielding the same output and the mutually conjugate coefficient vectors. The latter form is more common and the former was used in the original CLMS paper [26], in this paper, we will use the first form.

III. Augmented ESN

A. Standard Complex ESN with a Linear Output Mapping

Fig. 1 shows the architecture of a standard ESN, which is composed of K external input neurons, L readout neurons, and N internal units. Without loss of generality, we shall address ESNs with one readout neuron ($L = 1$), as this facilitates the nonlinear adaptive filtering setting within the ESN architecture. The input and internal weights are stored, respectively, in the $(N \times K)$ and $(N \times N)$ weight matrices $\mathbf{W}_{ip}$, $\mathbf{W}_{in}$, vector $\mathbf{w}_b$ comprises the feedback weights connecting the readout neuron and the internal units, vector $\mathbf{x}(k)$ is the $(N \times 1)$ internal state vector, $\mathbf{u}(k)$ represents the $(K \times 1)$ input vector, and $y(k)$ is the overall output. The network state at time instant k, denoted by $\mathbf{q}(k)$, is a concatenation of the input $\mathbf{u}(k)$, internal state $\mathbf{x}(k)$, and the delayed output $y(k-1)$

$$\mathbf{q}(k) = \big[u(k), \ldots, u(k-K+1), x_1(k), \ldots, x_N(k), y(k-1)\big]^T \tag{10}$$

whereas the internal unit dynamics are described by [12]

$$\mathbf{x}(k) = \mathbf{f}\big(\mathbf{W}_{ip}\mathbf{u}(k) + \mathbf{W}_{in}\mathbf{x}(k-1) + \mathbf{w}_b y(k-1)\big) \tag{11}$$

where $\mathbf{f}(\cdot)$ is a vector-valued nonlinear activation function of the neurons within the reservoir.

The echo state property is provided by randomly choosing an internal weight matrix $\mathbf{W}_{in}$ and performing scaling to make the spectral radius $\rho(\mathbf{W}_{in}) < 1$, thus ensuring that the network is stable, the input and feedback weights can be initialized arbitrarily [12]. For an ESN with a linear output mapping, the output $y(k)$ is given by

$$y(k) = \mathbf{q}^T(k)\mathbf{w}(k) \tag{12}$$

where $\mathbf{w}(k)$ is the weight vector corresponding to the output layer. Its update can be performed, e.g., based on the CLMS algorithm, given by [26]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\mathbf{q}^*(k). \tag{13}$$

B. Augmented Complex ESN with a Linear Readout Neuron

Based on the widely linear model given in Section II, we shall now derive the augmented widely linear stochastic gradient algorithm for the training of complex ESNs, thus making them suitable for processing general complex-valued signals (both circular and noncircular). To this end, we introduce the augmented network state $\mathbf{q}^a(k)$ as[4]

$$\begin{aligned}\mathbf{q}^a(k) \;=\; \big[&u(k), \ldots, u(k-K+1), x_1^a(k), \ldots, x_N^a(k),\\ &y(k-1), u^*(k), \ldots, u^*(k-K+1)\big]^T.\end{aligned} \tag{14}$$

Since the input weights of the ESN stored in matrix $\mathbf{W}_{ip}$ are randomly chosen prior to training, we can use another matrix $\mathbf{W}_{ip}^a$ to initialize the weights associated with the conjugate input vector $\mathbf{u}^*(k)$. The internal state transition within the augmented ESN is therefore described by

$$\begin{aligned}\mathbf{x}^a(k) \;=\; &\mathbf{f}\big(\mathbf{W}_{ip}\mathbf{u}(k) + \mathbf{W}_{in}\mathbf{x}^a(k-1)\\ &+\mathbf{w}_b y(k-1) + \mathbf{W}_{ip}^a\mathbf{u}^*(k)\big).\end{aligned} \tag{15}$$

[4]This augmented state is not a straightforward application of the widely linear model in (7) and is specific to ESNs.

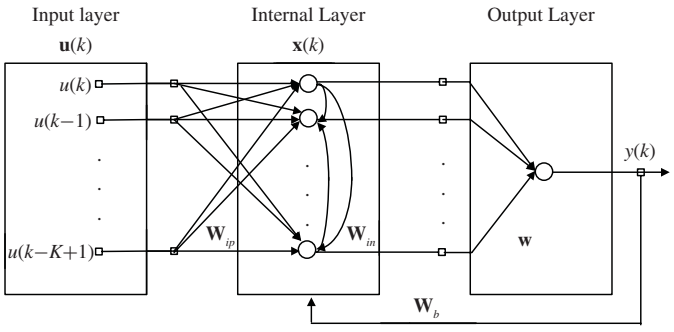

Fig. 1. Architecture of an ESN.

Due to the specific properties of the ESN architecture, the output $y(k)$ of the augmented ESN with linear output mapping is governed by an asymmetric version of the widely linear model in (7) to yield

$$y(k) = \mathbf{v}^T(k)\mathbf{h}(k) + \mathbf{u}^H(k)\mathbf{g}(k) \tag{16}$$

where $\mathbf{v}(k)$ is defined as $\mathbf{v}(k) = \big[u(k), \ldots, u(k-K+1), x_1^a(k), \ldots, x_N^a(k), y(k-1)\big]^T$, and $\mathbf{h}(k)$ and $\mathbf{g}(k)$ denote, respectively, the conventional and conjugate output weight vectors.

Note that the ESN has a local feedback (from the output to the internal state) and thus, unlike standard feedback structures [13], [28], the output within the state vector (14) of the ESN does not require augmentation with its conjugate. Therefore, due to local feedback, the conjugate weight vector $\mathbf{g}(k)$ is only associated with the conjugate input signal. The standard and conjugate weight vectors are thus of different dimensions, but as with all widely linear models, the conjugate weight vector $\mathbf{g}(k) = \mathbf{0}$ for a circular input signal.

C. Training of Augmented ESNs

The ESN is trained based on the cost function

$$\mathcal{J}(k) = \frac{1}{2}|e(k)|^2 = \frac{1}{2}e(k)e^*(k) \tag{17}$$

where $e(k)$ is the instantaneous output error $e(k) = d(k) - y(k)$, and $d(k)$ is the desired (teaching) signal. The update of the conjugate weight vector $\mathbf{g}(k)$ in (16) is given by

$$\mathbf{g}(k+1) = \mathbf{g}(k) - \mu\nabla_{\mathbf{g}}\mathcal{J}(k) \tag{18}$$

where μ is the learning rate. Note that $\mathcal{J}(k)$ is a real-valued function dependent on both output errors $e(k)$ and $e^*(k)$. It can be shown that the maximum change in the cost function on the error surface occurs in the direction of the conjugate gradient $\partial\mathcal{J}(k)/\partial\mathbf{g}^*(k)$, i.e., [29]–[31]

$$\nabla_{\mathbf{g}}\mathcal{J}(k) = \frac{\partial\mathcal{J}(k)}{\partial\mathbf{g}^*(k)}. \tag{19}$$

Expanding the term $\partial J(k)/\partial\mathbf{g}^*(k)$ gives

$$\nabla_{\mathbf{g}}\mathcal{J}(k) = \frac{1}{2}\left[e(k)\frac{\partial e^*(k)}{\partial\mathbf{g}^*(k)} + e^*(k)\frac{\partial e(k)}{\partial\mathbf{g}^*(k)}\right]. \tag{20}$$

Since

$$e^*(k) = d^*(k) - \mathbf{v}^H(k)\mathbf{h}^*(k) - \mathbf{u}^T(k)\mathbf{g}^*(k) \tag{21}$$

and $\partial e(k)/\partial \mathbf{g}^*(k) = \mathbf{0}$, we obtain

$$\nabla_{\mathbf{g}}\mathcal{J}(k) = -\frac{1}{2}e(k)\mathbf{u}(k) \tag{22}$$

giving the update of the conjugate weight vector $\mathbf{g}(k)$ in the form[5]

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu e(k)\mathbf{u}(k). \tag{23}$$

In a similar way, for the update of the conventional weight vector $\mathbf{h}(k)$, we have

$$e(k) = d(k) - \mathbf{v}^T(k)\mathbf{h}(k) - \mathbf{u}^H(k)\mathbf{g}(k) \tag{24}$$

with the gradient

$$\nabla_{\mathbf{h}}\mathcal{J}(k) = \frac{\partial J(k)}{\partial \mathbf{h}^*(k)} = \frac{1}{2}\left[e(k)\frac{\partial e^*(k)}{\partial \mathbf{h}^*(k)} + e^*(k)\frac{\partial e(k)}{\partial \mathbf{h}^*(k)}\right] \tag{25}$$

giving

$$\nabla_{\mathbf{h}}\mathcal{J}(k) = -\frac{1}{2}e(k)\mathbf{v}^*(k) \tag{26}$$

and the update

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu e(k)\mathbf{v}^*(k). \tag{27}$$

We can express the updates in (23) and (27) in a compact vector form, by defining the augmented weight vector $\mathbf{w}^a(k)$ as

$$\mathbf{w}^a(k) = \left[\mathbf{h}^T(k), \mathbf{g}^T(k)\right]^T \tag{28}$$

to give the update of ACLMS algorithm for ESNs

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \mu e(k)\mathbf{q}^{a^*}(k) \tag{29}$$

where $\mathbf{q}^a(k)$ is the augmented network state in (14).

Owing to the use of the widely linear model and augmented complex statistics, the augmented ESN has clear theoretical advantages over the standard ESN for the processing of noncircular complex signals. However, due to the sparse nature of the connectivity within the reservoir, the linear output mapping may not be powerful enough for efficient modeling of signals with large nonlinear dynamics. To this end, in the next section, we introduce a nonlinear readout neuron equipped with a trainable amplitude of nonlinearity.

D. Complex ESN with a Nonlinear Readout Neuron

The output $y(k)$ of the standard ESN with a nonlinear output layer is given by [12]

$$y(k) = \Phi(net(k)) = \Phi\left(\mathbf{q}^T(k)\mathbf{w}(k)\right) \tag{30}$$

where $\Phi(\cdot)$ is the output activation function, and the weight vector $\mathbf{w}(k)$ is updated by minimizing the cost function $\mathcal{J}(k)$, given in (17), to give

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\nabla_{\mathbf{w}}\mathcal{J}(k). \tag{31}$$

[5]The factor $1/2$ in (23) has been incorporated into the learning rate μ.

Similar as in the previous section, we can obtain the weight update of the complex nonlinear gradient descent (CNGD) algorithm [13], [32] for ESNs as[6] (factor $1/2$ is absorbed in μ)

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\Phi'^*\big(net(k)\big)\mathbf{q}^*(k). \tag{32}$$

To derive the corresponding augmented CNGD (ACNGD) algorithm for the nonlinear output layer of the augmented ESN, recall that the output $y(k)$ is given by

$$y(k) = \Phi(net(k)) = \Phi\left(\mathbf{v}^T(k)\mathbf{h}(k) + \mathbf{u}^H(k)\mathbf{g}(k)\right). \tag{33}$$

The update of the conjugate weight vector $\mathbf{g}(k)$ in (33) is given by

$$\mathbf{g}(k+1) = \mathbf{g}(k) - \mu\nabla_{\mathbf{g}}\mathcal{J}(k). \tag{34}$$

From (20), after setting $\partial e(k)/\partial \mathbf{g}^*(k) = \mathbf{0}$, we have

$$\nabla_{\mathbf{g}}\mathcal{J}(k) = \frac{1}{2}e(k)\frac{\partial e^*(k)}{\partial \mathbf{g}^*(k)}. \tag{35}$$

Since $e^*(k) = d^*(k) - \Phi^*(net(k))$, and for the complex transcendental functions $\partial\Phi^*(net(k))/\partial\mathbf{g}^*(k) = (\partial\Phi(net(k))/\partial\mathbf{g}(k))^*$, using the chain rule, we arrive at

$$\nabla_{\mathbf{g}}\mathcal{J}(k) = -\frac{1}{2}e(k)\Phi'^*\big(net(k)\big)\mathbf{u}(k) \tag{36}$$

giving the update for the weight vector $\mathbf{g}(k)$ in the form

$$\mathbf{g}(k+1) = \mathbf{g}(k) + \mu e(k)\Phi'^*\big(net(k)\big)\mathbf{u}(k). \tag{37}$$

In a similar way, the update of the weight vector $\mathbf{h}$(k) becomes

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \mu e(k)\Phi'^*\big(net(k)\big)\mathbf{v}^*(k) \tag{38}$$

and the update of the augmented weight vector

$$\mathbf{w}^a(k+1) = \mathbf{w}^a(k) + \mu e(k)\Phi'^*\big(net(k)\big)\mathbf{q}^{a^*}(k). \tag{39}$$

E. ESNs with an Adaptive Amplitude of Output Nonlinearity

The nonlinear output layer has been introduced into ESNs to provide a sufficient degree of nonlinearity for enhanced modeling, however, this does not automatically guarantee optimal modeling, as some parameters, such as the amplitude of the nonlinear readout neuron, need to be chosen empirically [33]. To this end, we now introduce a gradient adaptive amplitude into the output nonlinearity of ESNs. In this case, the activation function can be expressed as

$$y(k) = \Phi\big(net(k)\big) = \lambda(k)\overline{\Phi}\big(net(k)\big) \tag{40}$$

where the real-valued $\lambda(k) \geqslant 0$ denotes the amplitude of the nonlinearity $\Phi(net(k))$ and $\overline{\Phi}(net(k))$ is the unit amplitude activation function, for which $\lambda(k) = 1$. In the stochastic gradient setting, the parameter λ can be made gradient adaptive as [19], [33]

$$\lambda(k+1) = \lambda(k) - \eta\nabla\mathcal{J}_\lambda(k) \tag{41}$$

[6]For most standard nonlinear activation functions in $\mathbb{C}$ (transcendental functions such as *tanh*), we have $(\Phi^*)' = \Phi'^*$.

where η is the step size of the algorithm. The gradient

$$\nabla_\lambda \mathcal{J}(k) = \frac{\partial \mathcal{J}(k)}{\partial \lambda(k)} = \frac{1}{2}\frac{\partial\left[e(k)e^*(k)\right]}{\partial \lambda(k)} = \frac{1}{2}\left[e^*(k)\frac{\partial e(k)}{\partial \lambda(k)} + e(k)\frac{\partial e^*(k)}{\partial \lambda(k)}\right] \quad (42)$$

and since $\lambda(k)$ is real-valued, $\partial e^*(k)/\partial\lambda(k) = (\partial e(k)/\partial\lambda(k))^*$, giving

$$\frac{\partial e(k)}{\partial \lambda(k)} = -\overline{\Phi}(net(k)). \quad (43)$$

The update for the amplitude of nonlinearity within adaptive amplitude CNGD (AACNGD) algorithm is therefore given by

$$\lambda(k+1) = \lambda(k) + \frac{\eta}{2}\left[e^*(k)\overline{\Phi}(net(k)) + e(k)\overline{\Phi}^*(net(k))\right] \quad (44)$$

and applies to both standard (CGND) and ACNGD learning algorithms, for more detail, see [13].

F. Merits of Nonlinearity and Widely Linear Model

To illustrate the effect of nonlinearity and the widely linear model, we generated a circular doubly white noise $n(k)$, with zero mean and unit variance, and passed it through a complex-valued tanh nonlinearity, defined in (45), and the widely linear model given by $x(k) = \mathrm{WL}(n(k)) = 0.6n(k) + 0.8n^*(k)$. Fig. 2(a) and (b) shows that the application of the widely linear model does not change the Gaussian natures of the real and imaginary parts, it only alters the power ratio. Fig. 2(c) and (d) shows that the application of tanh nonlinearity alters the character of distribution, which cannot be achieved by using the widely linear model. It is therefore natural to combine the widely linear model and nonlinear processing in order to deal simultaneously with various aspects of the nature of the data. By introducing the adaptive amplitude of nonlinearity, we have an additional degree of freedom, allowing the nonlinear function to operate in a quasi-linear range if so required by the nature of the data.

IV. Simulations

To verify the potential of the proposed augmented ESNs compared to standard complex ESNs, we performed simulations on both benchmark synthetic proper and improper signals, and for noncircular real-world wind data.

For all signals, experiments were undertaken by averaging 200 independent simulation trials in the adaptive prediction setting. The nonlinearity at the nonlinear output layer of the ESNs was chosen to be the complex tanh function

$$\Phi(x) = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}} \quad (45)$$

with slope $\beta = 1$. Ten neurons were used in the hidden layer, with the internal connection weights having 5% degree of connectivity. The input tap length was $K = 1$, with no bias input. The values of the randomly selected input as well as internal and feedback weights $\mathbf{W}_{ip}$, $\mathbf{W}_{in}$, and $\mathbf{w}_b$ were taken from a uniform distribution in the range $[-1, +1]$, and the spectral radius $\rho(\mathbf{W}_{in})$ was set to be 0.8. The learning rate was

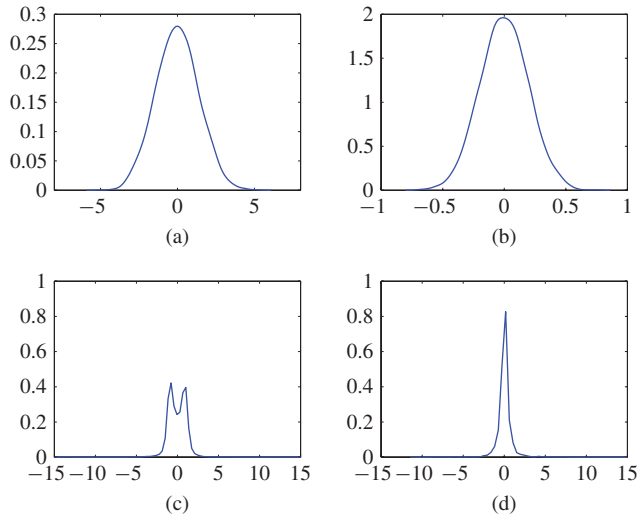

Fig. 2. Probability density functions obtained from a doubly white circular noise after applying the widely linear model and tanh nonlinearity. (a) Real part of WL ($n(k)$). (b) Imaginary part of WL ($n(k)$). (c) Real part of tanh ($n(k)$). (d) Imaginary part of tanh ($n(k)$).

$\mu = 0.005$ for all the learning algorithms considered, with the initial amplitude for the AACNGD algorithm $\lambda(0) = 1$ and the step size of the adaptive amplitude update $\eta = 0.2$.

The benchmark circular signal was a stable linear autoregressive (AR)(4) process, given by [2]

$$r(k) = 1.79r(k-1) - 1.85r(k-1) + 1.27r(k-3) - 0.41r(k-4) + n(k) \quad (46)$$

driven by complex-valued doubly circular white Gaussian noise $n(k)$ with zero mean and unit variance.

The benchmark noncircular signal was a complex AR moving-average (ARMA) process, whose transfer function was a combination of the MA model in [34] and the stable AR model in (46), given by

$$r(k) = 1.79r(k-1) - 1.85r(k-2) + 1.27r(k-3) - 0.41r(k-4) + 0.2r(k-5) + 2n(k) + 0.5n^*(k) + n(k-1) + 0.9n^*(k-1) \quad (47)$$

with

$$E\{n(k-i)n^*(k-j)\} = \delta(i-j)$$
$$E\{n(k-i)n(k-j)\} = \mathrm{C}\delta(i-j) \quad (48)$$

where $n(k)$ is the complex-valued doubly circular white Gaussian noise[7] and $\mathrm{C} = 0.95$ [34].

The nonlinear and noncircular chaotic Ikeda map signal is given by [35]

$$x(k+1) = 1 + u\,(x(k)cos[t(k)] - y(k)sin[t(k)])$$
$$y(k+1) = u\,(x(k)sin[t(k)] + y(k)cos[t(k)]) \quad (49)$$

where $u = 0.9$ and $t(k) = 0.4 - 6/(1 + x^2(k) + y^2(k))$.

[7]Double whiteness implies the uncorrelated real and imaginary channels, for circular white Gaussian noise, $n = n_r + \jmath n_i$, $\sigma^2_{n_r} = \sigma^2_{n_i}$, whereas for noncircular data, $\sigma^2_{n_r} > \sigma^2_{n_i}$, where $\sigma^2_{n_r}$ and $\sigma^2_{n_i}$ are, respectively, the powers of the real and imaginary parts.

TABLE I

COMPARISON OF DEGREES OF NONCIRCULARITY s FOR THE VARIOUS CLASSES OF SIGNALS

	Circular AR(4) (46)	Noncircular ARMA (47)	Ikeda map (49)	Wind (*low*)	Wind (*medium*)	Wind (*high*)
s	0.0016	0.9429	0.8936	0.1583	0.4305	0.8117

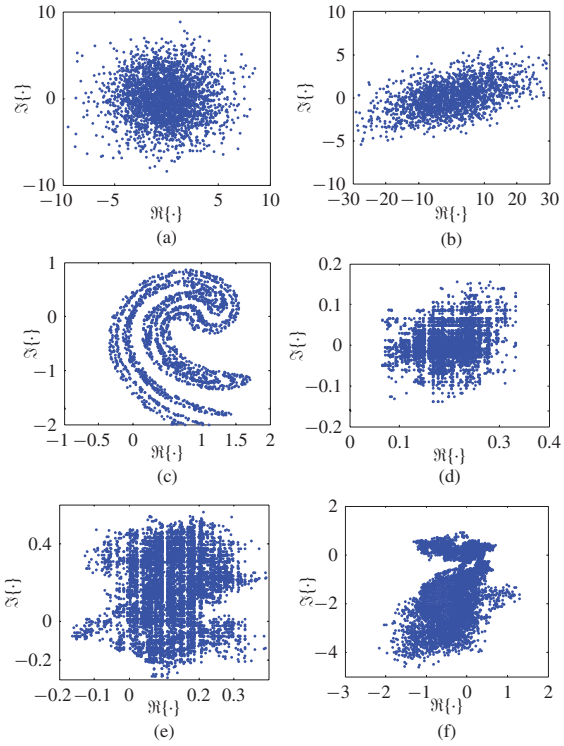

Fig. 3. Geometric view of circularity via "real-imaginary" scatter plots. (a) Circular AR(4) process (46). (b) Noncircular ARMA process (47). (c) Noncircular Ikeda map (49). (d) Wind (*low*) signal. (e) Wind (*medium*) signal. (f) Wind (*high*) signal.

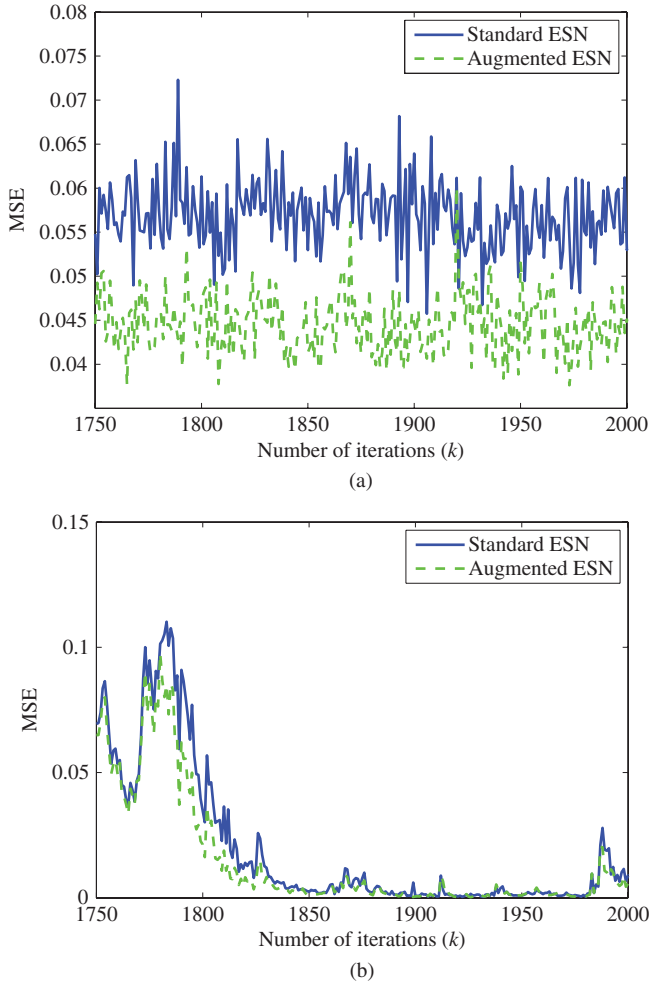

Fig. 4. MSEs of one-step-ahead prediction for augmented and standard ESN trained by the AACNGD algorithm. (a) Noncircular Ikeda map (49). (b) Noncircular and nonstationary wind (*high*) signal.

The wind data was sampled at 50 Hz and was collected in an urban environment over a one-day period [36], and was represented as a vector of speed and direction in the North–East coordinate system. The wind signal was made complex through combining the wind speed v and direction φ to form a complex signal $\boldsymbol{v} = \{ve^{j\varphi}\}$. Based on the changes in the wind intensity, the noncircular wind data were identified as regions of *low*, *medium*, and *high* dynamics. We here perform simultaneous prediction of wind speed and direction, the results corresponding to wind power can be found in [37] and [38].

Fig. 3 shows the scatter plots of the complex signals considered in simulations. Observe the circular symmetry (rotation invariance) for the AR(4) signal (46) and the noncircularity of the ARMA model (47), Ikeda map (49), and wind signals. For a quantitative measurement of the degree of noncircularity of a complex vector $\mathbf{z}$, we used the index s given by [39]

$$s = 1 - \det(\mathrm{C}_{\mathbf{z}^a\mathbf{z}^a})\det^{-2}(\mathrm{C}_{\mathbf{zz}}) \tag{50}$$

where $\det(\cdot)$ denotes the matrix determinant operator, the degree of noncircularity s is normalized to within [0, 1], with the value of 0 indicating perfect circularity. Table I illustrates the degrees of noncircularity s for the various classes of signals. Observe the excellent match between the measure of noncircularity in Table I and the scatter plot descriptions in Fig. 3, for instance, among the wind segments, the wind (*low*) region was least noncircular, whereas the wind (*high*) region exhibited strong noncircularity.

The standard prediction gain $R_p \triangleq 10\log_{10}\left(\sigma_x^2/\hat{\sigma}_e^2\right)[dB]$ was employed to assess the performance, where σ_x^2 and $\hat{\sigma}_e^2$ denote, respectively, the variance of the input signal $x(k)$ and the forward prediction error $e(k)$. Table II compares averaged prediction gains R_p (dB) and their standard deviations over 200 independent trials for the standard and augmented ESNs trained by CLMS, CNGD, and AACNGD algorithms, as well as a dual univariate ESN trained by the LMS algorithm[8] for the complex-valued signals considered. As expected, the dual univariate ESN, which treats the real and imaginary

[8]The dual univariate approach deals with complex-valued data by splitting the input signals into its real and imaginary parts and treating them as independent real-valued quantities [13]. In a dual univariate ESN, two independent reservoirs are generated to model the two components of the input.

TABLE II

COMPARISON OF PREDICTION GAINS R_p AND THEIR STANDARD DEVIATIONS (IN BRACKET) FOR THE VARIOUS CLASSES OF SIGNALS AVERAGED OVER 200 INDEPENDENT INITIALIZATIONS OF ESN

R_p [dB]	Circular AR(4)	Noncircular ARMA	Ikeda map	Wind (*low*)	Wind (*medium*)	Wind (*high*)
Dual univariate ESN (LMS)	4.1928 (0.5849)	2.8064 (0.3211)	−0.0271 (0.1846)	2.3519 (0.5882)	4.0615 (0.9153)	8.1162 (0.9124)
Standard ESN (CLMS)	4.6701 (0.6112)	3.5341 (0.3541)	2.1134 (0.3619)	2.4635 (0.4516)	4.7938 (0.7519)	9.5917 (0.9746)
Augmented ESN (ACLMS)	4.5489 (0.4792)	4.0396 (0.3624)	3.1967 (0.4258)	2.6417 (0.3620)	5.3029 (0.7781)	10.2063 (0.9537)
Standard ESN (CNGD)	4.6947 (0.5957)	3.6744 (0.3438)	2.1558 (0.3005)	2.6653 (0.4738)	5.0529 (0.7853)	9.9982 (1.0130)
Augmented ESN (Augmented CNGD)	4.5764 (0.4665)	4.1511 (0.3380)	3.2666 (0.5435)	2.8095 (0.3681)	5.7438 (0.8054)	10.7125 (0.9860)
Standard ESN (AACNGD)	6.6080 (0.5808)	5.0524 (0.5013)	2.4679 (0.5673)	4.0850 (0.3472)	6.3256 (0.7399)	11.7789 (0.9065)
Augmented ESN (Augmented AACNGD)	6.5357 (0.4970)	5.2484 (0.6149)	3.5912 (0.3053)	4.2571 (0.3499)	6.8286 (0.7968)	12.1900 (1.0272)

TABLE III

PERCENTAGE OF ENHANCED PERFORMANCE OF AUGMENTED ESN ALGORITHMS FOR COMPLEX NONCIRCULAR SIGNALS

	Noncircular ARMA	Ikeda map	Wind (*low*)	Wind (*medium*)	Wind (*high*)
Augmented ESN (ACLMS)	95.5%	98.5%	91.5%	94%	91%
Augmented ESN (Augmented CNGD)	95%	98%	90%	95%	94.5%
Augmented ESN (Augmented AACNGD)	93.5%	99%	89.5%	92.5%	92%

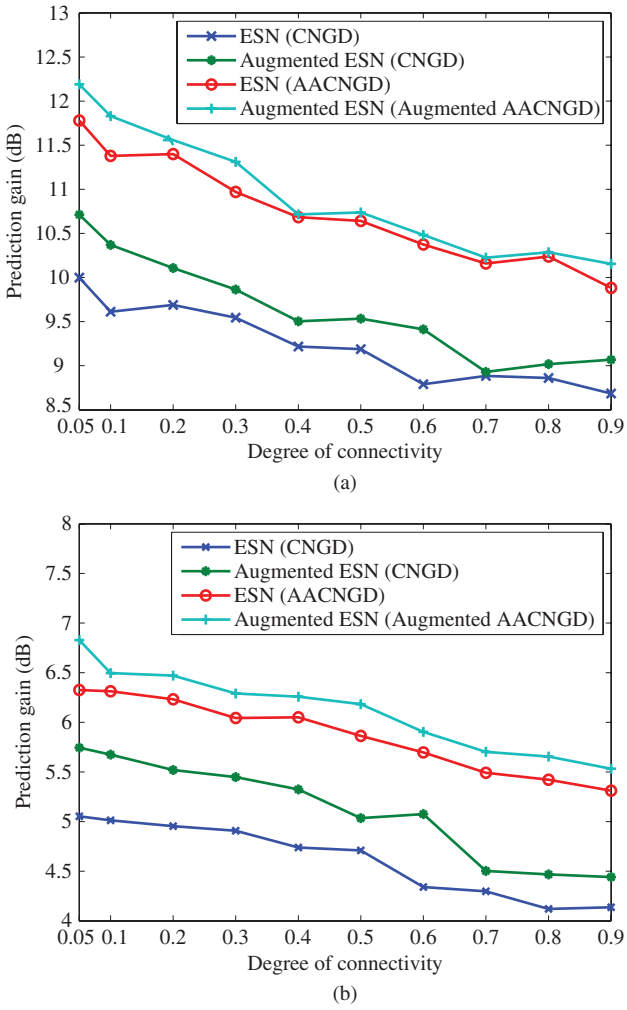

Fig. 5. Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of degrees of connectivity on one-step ahead prediction of the wind (*medium*) and (*high*) signals. (a) Wind (*high*). (b) Wind (*medium*).

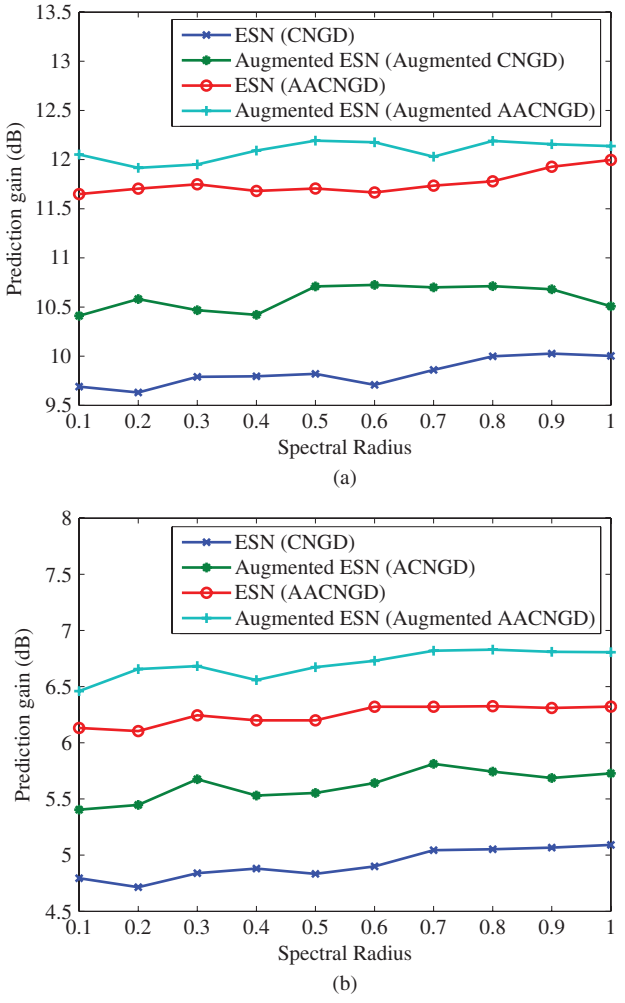

Fig. 6. Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of spectral radii on one-step ahead prediction of the noncircular wind (*medium*) and (*high*) signals. (a) Wind (*high*). (b) Wind (*medium*).

parts of complex-valued data as two independent channels, had the worst performance. For the circular AR(4) signal, the performance of the augmented complex ESN was similar to that of standard ESN. For the noncircular signals, there was a significant improvement in the prediction gain when the augmented ESN was employed. As desired, the advantage of the nonlinear output layer over the linear output mapping was more pronounced in the prediction of the nonlinear synthetic signal and nonlinear and nonstationary real-world wind signals. In practice, due to the randomly generated internal

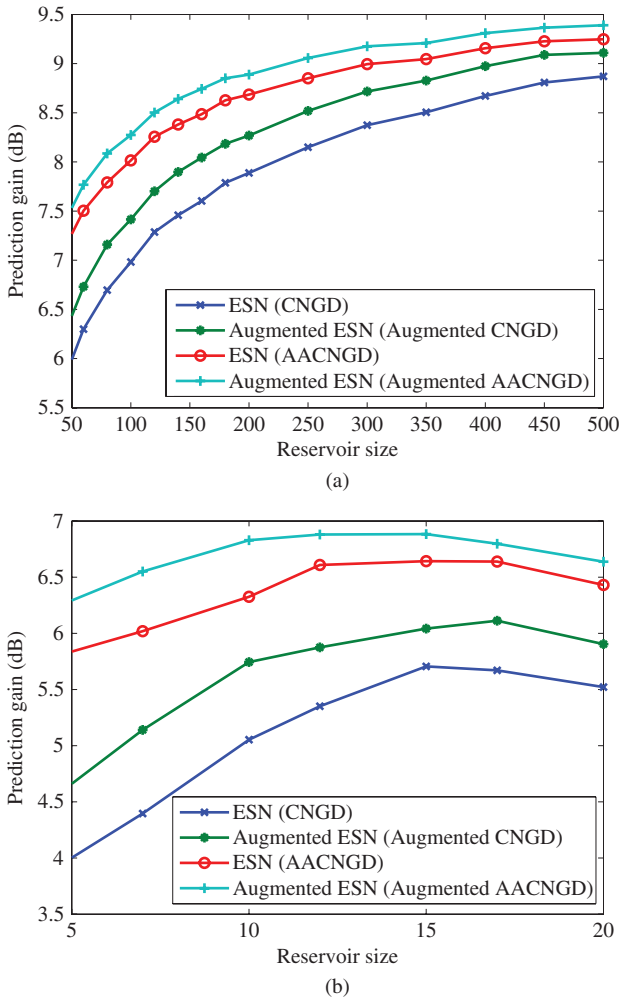

Fig. 7. Comparison of performances of standard and augmented ESNs trained by different algorithms, over a range of reservoir size on one-step-ahead prediction of the noncircular ARMA and wind (*medium*) signals. (a) Noncircular ARMA. (b) Wind (*medium*).

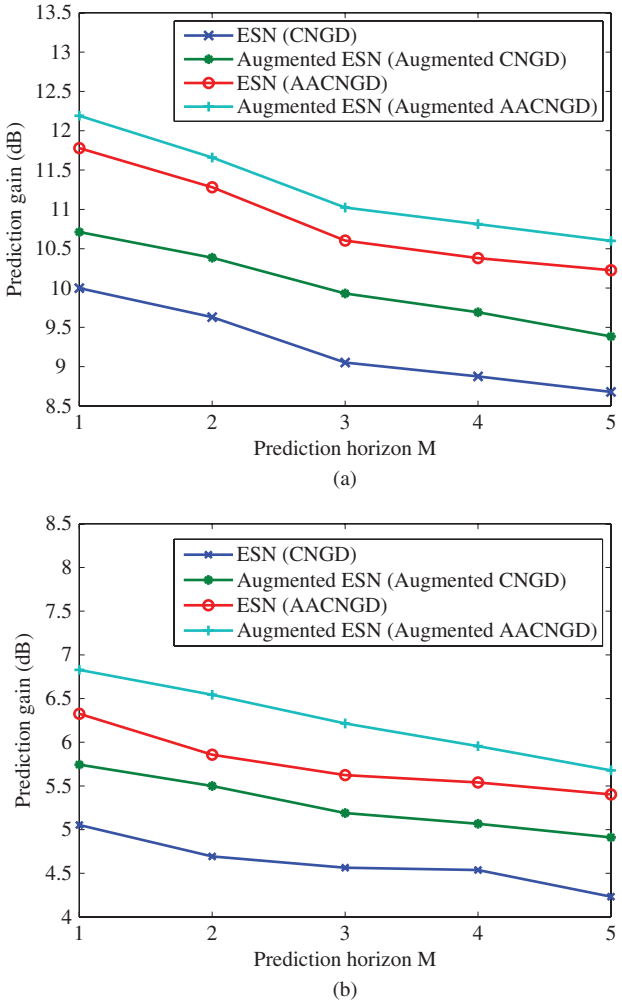

Fig. 8. Comparison of performances of standard and augmented ESNs trained by different algorithms on multiple-step-ahead prediction of the wind (*medium*) and (*high*) signals. (a) Wind (*high*). (b) Wind (*medium*).

reservoir within an ESN, the augmented ESN cannot guarantee enhanced performance over its standard version in every trial, however, as illustrated in Table III, on average, in more than 90% of the trials the widely linear algorithms outperformed the corresponding standard ones.

To further illustrate the advantage of using augmented complex statistics within complex-valued ESNs, we compared the MSEs of both the augmented and standard ESNs with adaptive amplitude of nonlinearity for the prediction of the complex-valued synthetic nonlinear and noncircular Ikeda map and the noncircular wind (*high*) signal. Fig. 4 shows that, in both cases, the augmented ESN with a nonlinear readout neuron trained by the augmented AACNGD outperformed its standard version.

We next investigated the influences of two parameters related to the generation of the internal layer, the degree of connectivity, and the spectral radius $\rho(\mathbf{W}_{in})$ on the performance of standard and augmented ESNs. Figs. 5 and 6 show that, in all the cases, for the prediction of real-world wind (*medium*) and (*high*) signals the augmented ESN trained by the augmented AACNGD algorithm achieved the best performance and that for both learning strategies it is desirable to keep a low degree of connectivity within the reservoir. This conforms to the ESN theory [12] that a small degree of connectivity can perform a relative decoupling of subnetworks with rich reservoir dynamics.

The size of the dynamical reservoir is another important parameter that influences the performance of ESNs, as it reflects their universal approximation ability. An ESN with a larger reservoir size can learn the signal dynamics with a higher accuracy [40], as shown in Fig. 7(a) on one-step-ahead prediction of the noncircular ARMA process in (47). This, however, applies to stationary signals, whereas for fast-changing nonstationary processes, the larger reservoir caused saturation of internal neurons, resulting in performance degradation, as shown in Fig. 7(b) for the prediction of the nonstationary wind (*medium*) signal. Observe that, in all the cases, the augmented ESNs outperformed their standard counterparts.

In the final set of simulations, we considered multistep-ahead prediction of the noncircular and nonstationary wind (*medium*) and (*high*) data. Fig. 8 shows the prediction gains of ESNs for a prediction horizon $M = 1, 2, 3, 4$, and 5, in all the cases, the augmented ESN with adaptive amplitude of nonlinearity achieved the best performance.

V. Conclusion

An augmented complex ESN has been introduced for nonlinear adaptive filtering of the generality of complex-valued signals. The proposed ESN has been derived based on augmented complex statistics, thus making it suitable for both second-order circular and noncircular signals. For generalities, a nonlinear output layer has been introduced, and to deal with signals with large dynamics, an adaptive amplitude has been introduced into the output layer of the augmented ESN. The proposed augmented ESNs have been shown to exhibit theoretical and practical advantages over their conventional counterparts. This has been verified through comprehensive simulations on both synthetic noncircular data and real-world wind measurements, and over a range of parameters.

Acknowledgment

The authors would like to thank Prof. Aihara's team at the Institute of Industrial Science, University of Tokyo, Tokyo, Japan, for providing wind data used in the simulations. They also acknowledge Gill Instruments, Hampshire, U.K., for providing their ultrasonic anemometers for the wind recordings.

References

[1] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Contr. Signals, Syst.*, vol. 2, no. 4, pp. 303–314, 1989.

[2] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. New York: Wiley, 2001.

[3] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Mar. 1990.

[4] C. H. Lee and C. C. Teng, "Identification and control of dynamic systems using recurrent fuzzy neural networks," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 4, pp. 349–366, Aug. 2000.

[5] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 526–535, Feb. 1995.

[6] D. P. Mandic and J. A. Chambers, "Toward an optimal PRNN-based nonlinear predictor," *IEEE Trans. Neural Netw.*, vol. 10, no. 6, pp. 1435–1442, Nov. 1999.

[7] S. A. Billings and C. F. Fung, "Recurrent radial basis function networks for adaptive noise cancellation," *Neural Netw.*, vol. 8, no. 2, pp. 273–290, 1995.

[8] C.-M. Lin, L.-Y. Chen, and D. S. Yeung, "Adaptive filter design using recurrent cerebellar model articulation controller," *IEEE Trans. Neural Netw.*, vol. 21, no. 7, pp. 1149–1157, Jul. 2010.

[9] J. L. Elman, "Finding structure in time," *Cognitive Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.

[10] W. Liu, I. Park, and J. C. Príncipe, "An information theoretic approach of designing sparse kernel adpative filters," *IEEE Trans. Neural Netw.*, vol. 20, no. 12, pp. 1950–1961, Dec. 2009.

[11] 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.

[12] H. Jaeger, "The echo state approach to analyzing and training neural networks," German Nat. Res. Inst. Inform. Technol., Sankt Augustin, Germany, Rep. 148, 2002.

[13] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York: Wiley, 2009.

[14] A. Hirose, *Complex-Valued Neural Networks*. New York: Springer-Verlag, 2006.

[15] I. N. Aizenberg, N. N. Aizenberg, and J. P. L. Vandewalle, *Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications*. Norwell, MA: Kluwer Academic, 2000.

[16] S. Jankowski, A. Lozowski, and J. M. Zurada, "Complex-valued multistate neural associative memory," *IEEE Trans. Neural Netw.*, vol. 7, no. 6, pp. 1491–1496, Nov. 1996.

[17] T. Nitta, "An extension of the back-propagation algorithm to complex numbers," *Neural Netw.*, vol. 10, no. 8, pp. 1391–1415, Nov. 1997.

[18] S. Seth, M. C. Ozturk, and J. C. Príncipe, "Signal processing with echo state networks in the complex domain," in *Proc. IEEE Workshop Mach. Learn. Signal Process.*, Thessaloniki, Greece, Aug. 2007, pp. 408–412.

[19] Y. Xia, D. P. Mandic, M. M. Van Hulle, and J. C. Príncipe, "A complex echo state network for nonlinear adaptive filtering," in *Proc. IEEE Workshop Mach. Learn. Signal Process.*, Cancun, Mexico, Oct. 2008, pp. 404–408.

[20] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[21] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[22] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and process," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 714–725, Mar. 2003.

[23] S. Javidi, M. Pedzisz, S. L. Goh, and D. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," in *Proc. 1st IARP Workshop Cognitive Inform. Process.*, 2008, pp. 54–57.

[24] S. L. Goh and D. P. Mandic, "An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks," *Neural Comput.*, vol. 19, no. 4, pp. 1039–1055, Apr. 2007.

[25] S. L. Goh and D. P. Mandic, "An augmented CRTRL for complex-valued recurrent neural networks," *Neural Netw.*, vol. 20, no. 10, pp. 1061–1066, 2007.

[26] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proc. IEEE*, vol. 63, no. 4, pp. 719–720, Apr. 1975.

[27] P. Schreier and L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[28] C. C. Took and D. P. Mandic, "Adaptive IIR filtering of noncircular complex signals," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4111–4118, Oct. 2009.

[29] D. H. Brandwood, "A complex gardient operator and its application in adaptive array theory," *IEE Proc. Commun., Radar Signal Process.*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[30] A. van den Bos, "Complex gradient and Hessian," *IEE Proc. Vis., Image Signal Process.*, vol. 141, no. 6, pp. 380–383, Dec. 1994.

[31] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," Dept. Elect. Comput. Eng., Univ. California, San Diego, Tech. Rep. ECE275A, 2006.

[32] G. M. Georgiou and C. Koutsougeras, "Complex domain backpropagations," *IEEE Trans. Circuits Syst. II*, vol. 39, no. 5, pp. 330–334, May 1992.

[33] E. Trentin, "Networks with trainable amplitude of activation functions," *Neural Netw.*, vol. 14, nos. 4–5, pp. 471–493, May 2001.

[34] J. Navarro-Moreno, "ARMA prediction of widely linear systems by using the innovations algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3061–3068, Jul. 2008.

[35] K. Aihara, *Applied Chaos and Applicable Chaos*. Tokyo, Japan: Science-Sha, 1994.

[36] D. P. Mandic, S. Javidi, S. L. Goh, A. Kuh, and K. Aihara, "Complex-valued prediction of wind profile using augmented complex statistics," *Renewable Energy*, vol. 34, no. 1, pp. 196–210, Jan. 2009.

[37] G. Sideratos and N. Hatziargyriou, "An advanced statistical method for wind power forecasting," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 258–265, Feb. 2007.

[38] R. J. Bessa, V. Miranda, J. C. Príncipe, A. Botterud, and J. Wang, "Information theoretic learning applied to wind power modelling," in *Proc. IEEE World Congr. Comput. Intell.*, Barcelona, Spain, Jul. 2010, pp. 2409–2416.

[39] P. J. Schreier, "The degree of impropriety (noncircularity) of complex random vectors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Las Vegas, NV, Apr. 2008, pp. 3909–3912.

[40] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and the "echo state network" approach," German Nat. Res. Inst. Inform. Technol., Sankt Augustin, Germany, Rep. 159, 2002.

Yili Xia (S'09) received the B.Eng. degree in information engineering from Southeast University, Nanjing, China, in 2006, and the M.Sc. (distinction) degree in communications and signal processing from the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., in 2007. He is currently pursuing the Ph.D. degree at the Imperial College London, where he is a Research Assistant.

His current research interests include linear and nonlinear adaptive filters and complex-valued statistical analysis.

Beth Jelfs received the M.Eng. degree in electronic and software engineering from the University of Leicester, Leicester, U.K. She is currently working toward the Ph.D. degree in electrical and electronic engineering at the Imperial College London, London, U.K.

She was with Marconi Optical Components, Northamptonshire, U.K., as a Test Technician in 2000. She is involved on a British Council Research Exchange Grant with the Technische Universität München, München, Germany, and the Max-Planck Institute for Dynamics and Self-Organization, Göttingen, Germany. Her current research interests include adaptive signal processing and signal modality characterization.

Ms. Jelfs was a recipient of the British Computer Society Prize for the top graduate from the University of Leicester in 2005.

Marc M. Van Hulle (SM'00) received the M.Sc. degree in electrotechnical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven (K. U. Leuven), Leuven, Belgium. He also received the B.Sc.Econ. and M.B.A. degrees. He received the Doctor Technices degree from Queen Margrethe II of Denmark in 2003, and an Honorary Doctoral degree from Brest State University, Brest, Belarus, in 2009.

He is currently a Full Professor at the K. U. Leuven Medical School, where he heads the Computational Neuroscience Group of the Laboratorium voor Neuro- en Psychofysiologie. In 1992, he was with the Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, Boston, as a Post-Doctoral Scientist. He has authored a monograph titled *Faithful Representations and Topographic Maps: From Distortion- to Information-Based Self-Organization* (John Wiley, 2000; also translated into Japanese), and has published 200 technical papers. His current research interests include computational neuroscience, neural networks, computer vision, data mining, and signal processing.

Dr. Van Hulle is an Executive Member of the IEEE Signal Processing Society, and an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS, *Computational Intelligence and Neuroscience*, and the *International Journal of Neural Systems*. He is a member of the program committees of several international machine learning and signal processing workshops and conferences. In 2009, he received the SWIFT prize of the King Baudouin Foundation of Belgium for his work on the Mind Speller, which received worldwide press coverage. In 2010, he received the Red Dot Design Award for the Mind Speller, which is one of the most prestigious design prizes in the world.

José C. Príncipe (M'83–SM'90–F'00) received the Honorary Doctorate degree from the Universita Mediterranea, Reggio di Calabria, Italy, the University of Maranhao, São Luis, Brazil, and Aalto University, Helsinki, Finland.

He is a Distinguished Professor of electrical and biomedical engineering at the University of Florida, Gainesville, where he teaches advanced signal processing and machine learning. He is the BellSouth Professor and Founding Director of the Computational Neuro-Engineering Laboratory, University of Florida. He has authored five books and published more than 200 book chapters and research papers in refereed journals, and has presented over 380 conference papers. He has directed 65 Ph.D. dissertations and 67 Master's theses. His current research interests include advanced signal processing and machine learning, brain–machine interfaces, and the modeling and applications of cognitive systems.

Dr. Príncipe is a Fellow of the American Institute for Medical and Biological Engineering. He is the recipient of the INNS Gabor Award, the IEEE Engineering in Medicine and Biology Society Career Achievement Award, and the IEEE Computational Intelligence Society Neural Network Pioneer Award. He is Editor-in-Chief of the IEEE REVIEWS ON BIOMEDICAL ENGINEERING, Past Editor-in-Chief of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, current Administrative Committee Member of the IEEE Computational Intelligence Society, the IEEE Biometrics Council, and the IEEE Biomedical Engineering Society, member of the Technical Committee on Machine Learning for Signal Processing of the IEEE Signal Processing Society, member of the Executive Committee of the International Neural Network Society, and past President of the International Neural Network Society. He is a former member of the Scientific Board of the Food and Drug Administration, and a member of the Advisory Board of the McKnight Brain Institute, University of Florida.

Danilo P. Mandic (M'99–SM'03) received the Ph.D. degree in nonlinear adaptive signal processing from Imperial College London, London, U.K., in 1999.

He is a Reader in signal processing at the Imperial College London. He has been a Guest Professor at Katholieke Universiteit Leuven, Leuven, Belgium, Tokyo University of Agriculture & Technology, Tokyo, Japan, and Westminster University, London, and a Frontier Researcher in RIKEN, Saitama, Japan. He has been working in the area of nonlinear adaptive signal processing and nonlinear dynamics. His publication record includes two research monographs titled *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability* (1st ed., Aug. 2001) and *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models* (1st ed., Wiley, Apr. 2009), an edited book titled *Signal Processing Techniques for Knowledge Extraction and Information Fusion* (Springer, 2008) and more than 200 research papers on signal and image processing.

Dr. Mandic is a member of the London Mathematical Society. He is a member of the IEEE Technical Committee on Machine Learning for Signal Processing, Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS, and the *International Journal of Mathematical Modelling and Algorithms*. He has produced award winning papers and products resulting from his collaboration with industry.