Influence of data preprocessing on prediction of complex valued load time series

Igor Krcmar, Petar Matic
Faculty of Electrical Engineering
University of Banja Luka
Banja Luka, Bosnia and Herzegovina
igor.krcmar@etfbl.net, petar.matic@etfbl.net

Slavisa Krunic
Siemens d.o.o. Banjaluka
Banja Luka, Bosnia and Hrzegovina
slavisa.krunic@siemens.com

Abstract— **Exploitation of modern power systems requires prediction of the electrical load time series for operation of power utilities and load estimates for market operation and system planning. Increase of energy produced from renewable sources and deregulation of electrical energy market makes load prediction more important nowadays. By its nature, electrical load time series are highly non linear and require modeling in the complex domain. Therefore, neural network based models, with fully complex activation functions, are appropriate choice for prediction of electrical load time series. However, their performance can be affected by input data preprocessing. Due to that cause, the paper analyses influence of data preprocessing on prediction of complex valued load time series. The analysis is performed on metered load data, that represents fifteen minutes average of active and reactive power, obtained from the medium voltage grid and with application of simple predictor structures, i.e. neural adaptive filters, applied to the one step ahead prediction tasks.**

Keywords- complex valued load; time series prediction; data preprocessing; neural adaptive filter

I. INTRODUCTION

Deregulation of energy market, as well as increase of amount of energy obtained from renewable sources, imposes electrical load prediction as a necessity, in everyday operation of the key players on the market [1]. Electrical load time series are highly non linear, non stationary, and complex valued. Therefore, neural network (NN) based models are appropriate solution for the task of electrical load prediction [2]-[4]. Further, in order to reflect complex nature of electrical load time series, NN should employ fully complex (FC) activation function (AF) at a neuron [5]-[7]. Usual choice for the FC AF, within an NN based model, are meromorphic functions, i.e. functions analytic everywhere, except on a discrete subset of the set of complex numbers, C [8]. Thus, FC AF can facilitate gradient descent learning [9],[8]. Performance of the NN based models is dominantly determined by the size and structure of the training set, learning algorithm, and structure of the model [10]. Operation of a gradient descent learning algorithm, which provides simple organization of the model weights update, is dominantly determined by the value of the learning rate parameter and persistency of excitation (PE). If an input signal is PE of low order, then optimal value of weights will not be reached through the learning process [11],[12]. Another problem, regarding the task of time series prediction, is non stationary character of time series at hand or existence of certain trend within the time series [13],[14]. These will result in change of optimal value of weights and might slow down the learning process, or decrease its overall performance. Therefore, input data should be processed prior they enter the learning process. Input data preprocessing should provide higher order PE, avoid saturation of neurons, and try to eliminate trends, if any is present in the time series. However, data preprocessing techniques should take into account the fact that electrical load signal is complex valued, so its phase angle, also, carries certain amount of information on the process of electrical energy consumption. Neural adaptive filters, due to their simple structure and gradient descent learning algorithms, are good tool for time series analysis [15],[7],[8]. Therefore, they will be applied in order to analyze different data preprocessing techniques.

The paper is organized as follows. The second Section provides an overview of AF, FC and real valued, as well as some of the standard data preprocessing techniques. The third Section gives principles of operation of neural adaptive filters, i.e. their structure and some gradient descent learning algorithms. The fourth Section contains experimental analysis of different data processing techniques. The experiments are performed as one step ahead prediction of the metered values of active and reactive power, obtained from the medium voltage grid. The conclusions are given in the fifth Section.

II. ACTIVATION FUNCTIONS AND DATA PREPROCESSING

Standard choice for a real valued AF at the neuron, within NN or neural adaptive filters are sigmoid functions. They have some nice properties, i.e. they are differentiable and bounded [16],[10]. As an example, we shall consider logistic AF, given by

$$\Phi(x) = \frac{1}{1+\exp(-\beta x)}, \tag{1}$$

where β denotes the slope of the AF, $\Phi(\bullet)$ denotes AF of the output neuron, and x is a variable. Real valued logistic AF, for different values of β is presented on Fig. 1. Absolute value of a complex valued logistic activation function, for $\beta = 1$, is presented on the Fig.2, while its phase angle is given on the

Fig.3. Having in mind that real valued logistic AF is bounded, the first type of data preprocessing is scaling of data so they can fit the range of the AF. Usual choice for data scaling is given by

$$x_s = p\frac{x - x_{\min}}{x_{\max} - x_{\min}} + q, \quad (2)$$

where x_s denotes new, i.e. scaled data, $x_{\min}$ is minimum value of the original time series, $x_{\max}$ is maximum value of the original

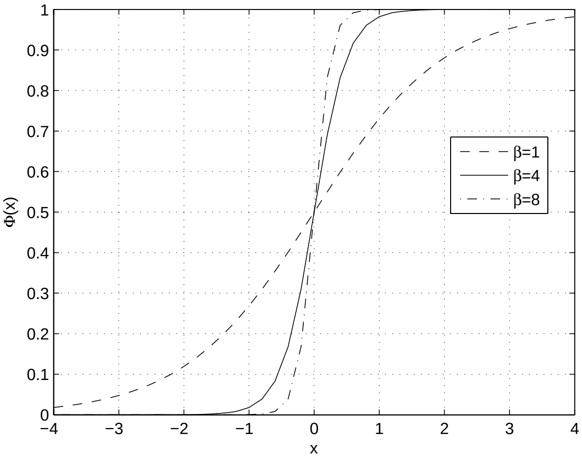

Figure 1. Real valued logistic activation function for different values of the slope parameter

time series, while real numbers p and q defines the range of new data in a way $x_s \in [q, q{+}p]$. With proper choice of p and q, one can achieve an adequate order of PE and avoid saturation of the neurons. Saturation of the neurons will slow down or stop the learning process [17]. From the Fig. 2 it is obvious that complex valued logistic function is not bounded, but with certain activation, neuron with this AF will go to saturation. To avoid this situation, input data should be preprocessed prior to the learning process. If one tries to apply (2) to solve the problem, he will face the following obstacles. Firstly, set of complex numbers is not an ordered set, thus $x_{\min}$ and $x_{\max}$ do not have meaning within the set of complex numbers. The second, complex logistic AF does not have clearly defined range, so data transformation can target two issues, i.e. the PE of an input signal and saturation of the neuron. The third issue, regarding the application of (2) on complex valued time series, is preservation of the phase angle. Thus, one can use signal scaling, given by the following equation

$$x_s = K\frac{x}{\max_x |x|}, \quad (3)$$

where |•| denotes the absolute value. Even though it is simple, transformation given by (3) preserves phase angle, and by proper choice of the real number K, the PE can be increased, thus improving the learning process. Removal of a trend, present in the time series, cannot be achieved neither by (2) nor by (3). A simple way to remove the trend from the time series is differentiation of the time series, i.e. generation of the new time series as follows

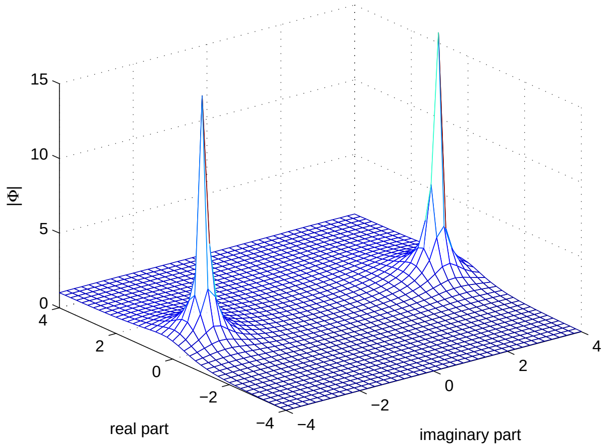

Figure 2. Absolute value of the complex valued logistic activation function for β = 1

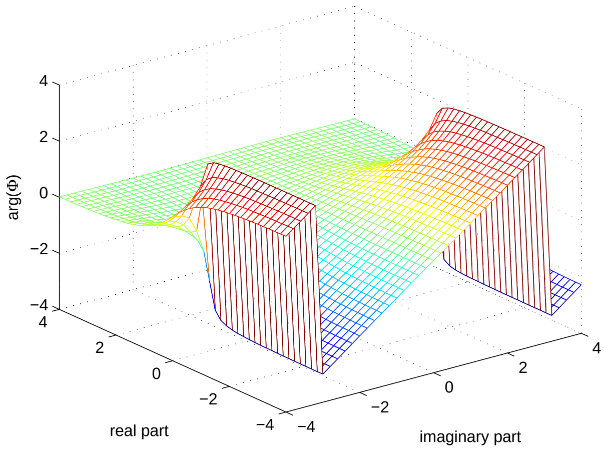

Figure 3. Phase angle of the complex valued logistic activation function for β = 1

$$x_d(k) = \Delta x(k) = x(k) - x(k-1), \quad (4)$$

where k denotes the discrete time instant, Δ denotes the differentiation operator, and x_d denotes new time series, the time series without the trend. Obviously, the new time series x_d should be modified by (3), before it enters the learning process.

III. Neural Adaptive Filters

The structure of a finite impulse response (FIR) complex valued neural adaptive filter is given on the Fig. 4. Equations that describe operation of the filter are as follow

$$y(k) = \Phi(net(k)), \quad (5)$$

$$net(k) = \mathbf{x}^T(k)\mathbf{w}(k), \tag{6}$$

where $y(k)$ denotes the filter output, , $net(k)$ is activation of the neuron, N is the length of filter tap inputs, $\mathbf{w}(k)=[w_1(k), w_2(k),\dots, w_N(k)]^T$ is the filter weight vector, $(\bullet)^T$ denotes vector transpose and $\mathbf{x}(k)=[x_1(k), x_2(k),\dots, x_N(k)]^T$ is the filter input

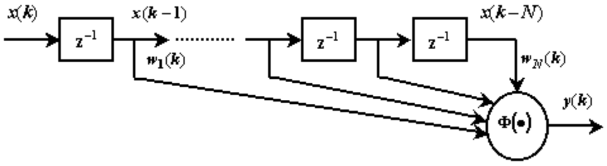

Figure 4. A neural adaptive complex valued FIR filter

vector, where $x_i(k)= x(k-i)$, $i=1,2,\dots,N$. For the filter, given on the Fig. 4, a stochastic gradient descent learning algorithm is described by the following equations [6],[15]

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\nabla_{\mathbf{w}} J(k), \tag{7}$$

$$J(k) = \frac{1}{2}\left|e(k)\right|^2, \tag{8}$$

$$e(k) = d(k) - y(k), \tag{9}$$

where μ denotes the learning rate, $J(k)$ is the cost function, $\nabla_{\mathbf{w}}(\bullet)$ denotes gradient of a scalar function with respect to the weight vector $\mathbf{w}$, $e(k)$ is the error at the output neuron and $d(k)$ is some desired, i.e. teaching signal. Computation of the gradient $\nabla_{\mathbf{w}} J(k)$ is not trivial, due to the fact that the cost function is not complex analytic [18], [19],[20]. In the case Φ is a meromorphic function [8], computation of the gradient of the cost function (8) gives [8], [19].

$$\begin{aligned}\nabla_{\mathbf{w}} J(k) &= \nabla_{\mathbf{w}_r} J(k) + j\nabla_{\mathbf{w}_i} J(k) - \mathbf{x}^*(k) \\ &= \left[\Phi'^*(k)e_r(k) + j\Phi'^*(k)e_i(k)\right] \\ &= -e(k)\mathbf{x}^*(k)\Phi'^*(k),\end{aligned} \tag{10}$$

where $j=\sqrt{-1}$, $\mathbf{w}_r=\Re(\mathbf{w})$, $\mathbf{w}_i=\Im(\mathbf{w})$, $e(k) = \Re(e(k)) + j\Im(e(k)) = e(k) + je_i(k)$, and where for convenience $\Phi'^*(net(k)) = \Phi'^*(k)$. Further, $(\bullet)'$ denotes the first derivative and $(\bullet)^*$ denotes the complex conjugate. Now, from (7) and (10) we have the weight update equation for the complex nonlinear gradient descent (CNGD) algorithm

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\Phi'^*(k)\mathbf{x}^*(k). \tag{11}$$

In a normalized stochastic gradient learning algorithm, the learning rate μ is not constant. The learning rate, which yields the normalized complex nonlinear gradient descent (NCNGD) algorithm has the learning rate $\mu_{NCNGD}(k) = \eta / (C + |\Phi'(k)|^2 \|\mathbf{x}(k)\|_2^2)$. It is optimal in the sense that it minimizes the value of the *a posteriori* error [6], [8], [15] $d(k) - \Phi(\mathbf{x}^T(k)\,\mathbf{w}(k+1))$. The NCNGD algorithm robust on the value of its design parameters [17], thus it is a natural choice when it comes to application of neural adaptive filters on the problem of hypothesis testing [7].

IV. Experimental Analysis

The experiments were carried out, as one step ahead signal prediction, in order to compare different data preprocessing techniques. The test complex valued electrical load signal is shown on the Fig. 5 and the Fig. 6.

The signal represents fifteen minutes average of active and reactive power, metered at the 10 kV feeder, in the Transformer station Banja Luka 2.The complex valued logistic AF was used within the experiments as nonlinearity at the neuron, with the slope β

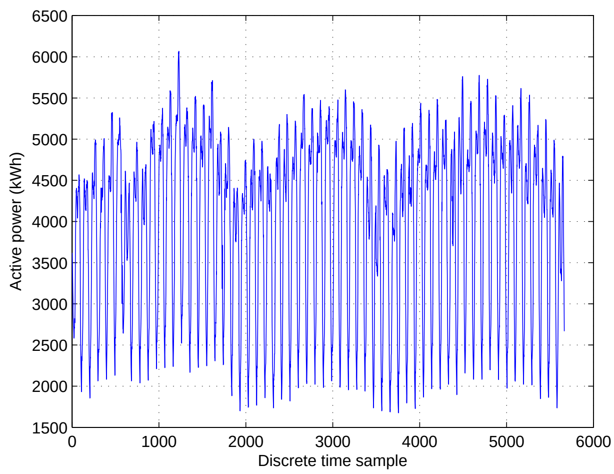

Figure 5. Real part of the test signal

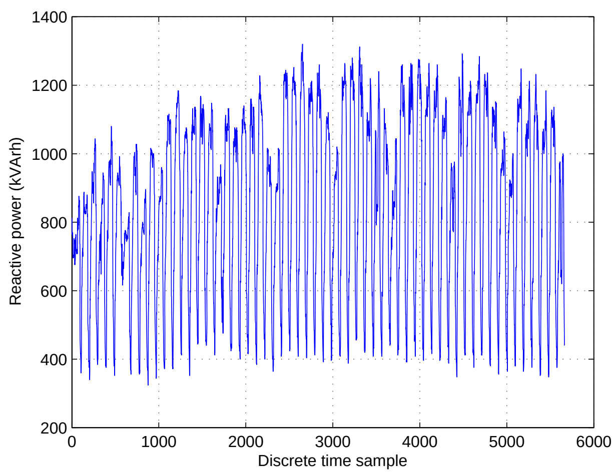

Figure 6. Imaginary part of the test signal

= 4. In all the experiments the NCNGD algorithm was applied, with the learning rate parameter μ=0.1 and C=0.1.

The performance measure was standard prediction gain $R_p=10\log_{10}(\sigma_y^2/\sigma_e^2)$, where σ_y^2 and σ_e^2 denote variances of the predicted signal and the output error, respectively. Value of the coefficient K was varied from 0.1 up to 4, with the step 0.1. Order of the neural adaptive filter was from the set N = {10, 20, 30, 40, 50}. In the first experiment, the test signal was scaled according to (3). Summary of the first experiment is given on the Fig. 7. In the second experiment, data preprocessing was performed according to (4) and (3). Results of the second experiment are summarized in the Fig. 8. From the Fig. 7 it is obvious that each performance curve has a unique maximum. The maximum is achieved for certain $K = K_{\max}$, and $1 < K_{max} < 1.5$.

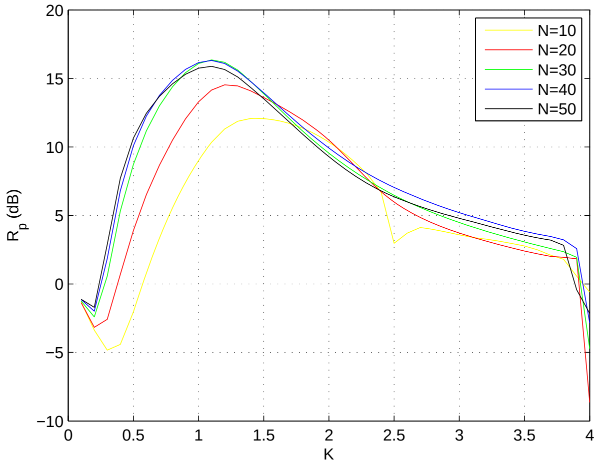

Figure 7. Results of the first experiment

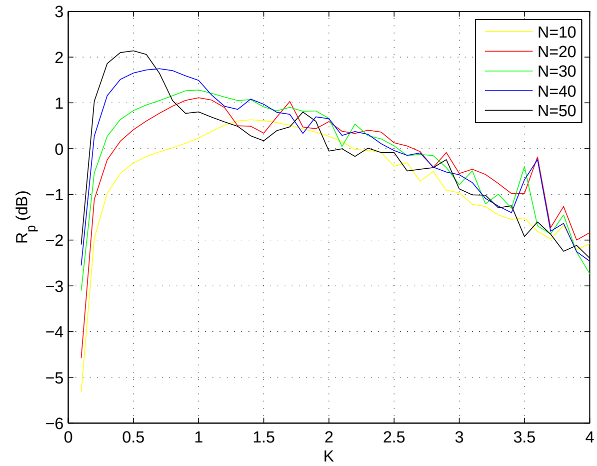

Figure 8. Results of the second experiment

For the $K < K_{max}$ input data does not have sufficient PE, therefore learning process cannot yield optimal value of the filter weights. In the case $K > K_{max}$ learning algorithm forces the output neuron to saturation, thus the learning process is too slow and cannot yield the optimal filter weights. Results of the second experiment reveal the same shape of the performance curves, as in the first experiment, i.e. the same line of reasoning holds. Significantly lower values of the performance indices in the second experiment, comparing to the results of the first experiment, means that data preprocessing according to (4) extracts certain amount of information, thus reducing capability of the learning process.

V. CONCLUSIONS

Performance analysis of some simple data preprocessing techniques, for application in one step ahead complex valued load prediction tasks, has been given. Data preprocessing is very important step in the design procedure of every NN based model for complex valued electrical load prediction. Neural adaptive FIR filters, with the NCNGD algorithm, have been employed in order to assess performance of these techniques. Simple scaling of an input signal, which preserves phase angle, has shown good performance. With one free parameter, it can improve, to some extent, PE of an input signal, while avoiding saturation of the neuron. Differentiation of an input signal, as a tool for the trend removal, has shown poor performance. Further research has to be done, in order to find adequate modification of the basic procedure.

REFERENCES

[1] P. Failey, "The unruly power grid," *IEEE Spectrum*, vol. 41, issue 8, pp. 22-27, Aug. 2004.

[2] S. Krunic, I. Krcmar, and N. Rajakovic, "An improved neural network application for short-term load forecasting in power systems," *Electric Machines and Power Systems*, vol. 28, pp. 703-721, 2000.

[3] A. Khotanzad, R. A. Rohani, T. L. Lu, A. Abaye, M. Davis, and D. J. Maratukulam,"ANNSTLF-A neural-network-based electric load forecasting system,"*IEEE Transactions on Neural Networks*, vol. 8, pp. 835-846, Jul. 1997.

[4] R. C. Bansal and J. C. Pandey, "Load forecasting using artificial intelligence techniques: a literature survey," *International Journal of Computer Applications in Technology*, vol. 22, no. 2/3, pp. 109-119, 2005.

[5] S. Krunic and I. R. Krcmar, "Comparative analysis of artificial neural network models for complex valued load prediction," in *Proceedings of 8th Conference on Electricity Distribution in Serbia, with regional participation (CIRED 2012)* , R-6.03, Vrnjacka Banja, 2012.

[6] Igor R. Krcmar, Petar S. Maric, and Milorad M. Bozic, "A Class of Neural Adaptive FIR Filters for Complex-Valued Load Prediction," *Proceedings of* 10th *Seminar on Neural Network Applications in Electrical Engineering (NEUREL-2010)*, pp. 37–39, 2010.

[7] S. Krunic, I. R. Krcmar, P. S. Maric, and M. M. Bozic, "Application of Neural Adaptive Filters for Analysis," in *Proceedings of 9th Symposium on Industrial Electronics (INDEL 2012)* , pp. 176-179, Banja Luka, 2012.

[8] D. P. Mandic and S. l. Goh, *Complex valued nonlinear adaptive filters: noncircularity, widely linear and neural models*, John Wiley & Sons, 2009.

[9] T. Kim and T. Adali, "Approximation by fully complex multilayer perceptrons," *Neural Computations*, vol. 15(7), pp. 1641-1666, 2003.

[10] S. Haykin, *Neural networks - A comprehensive foundation*. Prentice Hall, 1994.

[11] P. Stoica and S. Soderstrem, *System Identification*. John Wily and Sons,1987.

[12] L. Guo, L. Ljung, and G. J. Wang, "Necessary and sufficient conditions for stability of LMS," *IEEE Transactions on Automatic Control*, vol. 42, no. 6, pp. 761–770, 1997.

[13] S. L. Goh and D. P. Mandic, "Nonlinear adaptive prediction of complex valued non-stationary signals," *IEEE Transactions on Signal Processing*, vol. 53(5), pp. 1827-1836, 2005.

[14] S. L. Goh, M. Chen, D. H. Popovic, K. Aihara, D. Obradovic, and D. P. Mandic, "Complex-valued forecasting of wind profile," *Renewable Energy*, vol. 31(11), pp. 1733-1750, 2006.

[15] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, John Wiley & Sons, 2001.

[16] A. I. Hanna, I. R. Krcmar, and D. P. Mandic, "Perlustration of error surfaces for linear and nonlinear stochastic gradient descent algorithms," *Proceedings of* 6th *Seminar on Neural Network Applications in Electrical Engineering (NEUREL-2002)*, pp. 11-16, 2002.

[17] I. R. Krcmar, P. S. Maric, and M. M. Bozic, "Comparative Analysis of the DRNGD and the NNGD Algorithms in Complex Dom," in *Proceedings of 9th Symposium on Industrial Electronics (INDEL 2012)* , pp. 171-175, Banja Luka, 2012.
[18] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proceedings of the IEEE*, vol. 63(3), pp. 712-720, 1975.
[19] K. Kreutz-Delgado, *The complex gradient operator and the CR calculus*, Lecture supplement ECE275A, pp. 1-74, 2006.
[20] T. Needham, *Visual Complex Analysis*, Oxford University Press, 1997.