Comparison of Echo State Network and Extreme Learning Machine for PV Power Prediction

Iroshani Jayawardene, *Student Member, IEEE* and Ganesh K. Venayagamoorthy, *Senior Member, IEEE*
Real-Time Power and Intelligent Systems (RTPIS) Laboratory
The Holcombe Dept. of Electrical and Computer Engineering
Clemson University, SC 29634, USA
iroshani.jayawardene.2014@ieee.org and gkumar@ieee.org

***Abstract*— The increasing use of solar power as a source of electricity has introduced various challenges to the grid operator due to the high PV power variability. The energy management systems in electric utility control centers make several decisions at different time scales. In this paper, power output predictions of a large photovoltaic (PV) plant at eight different time instances, ranging from few seconds to a minute plus, is presented. The predictions are provided by two learning networks: an echo state network (ESN) and an extreme learning machine (ELM). The predictions are based on current solar irradiance, temperature and PV plant power output. A real-time study is performed using a real-time and actual weather profiles and a real-time simulation of a large PV plant. Typical ESN and ELM prediction results are compared under varying weather conditions.**

***Keywords*—Echo state network; extreme learning machine; PV; power prediction; real-time weather**

I. INTRODUCTION

Photovoltaic (PV) power generation is one of the major energy sources among renewable energy sources. The increasing use of PV power as a source of electricity has introduced various challenges to the grid operator. Unexpected large variations of PV plant output can cause dynamic and transient stability concerns which may lead to the collapse of a power system. This increases operating costs for the electricity system and potential risks to the reliability of electricity supply. Predicting variations of the PV power generation is a potential solution to overcome these challenges.

The power generated in a PV plant fluctuates with the weather changes. The output power produced by PV plants usually varies with solar irradiance. The irradiance is zero at night and starts to increase gradually during the day and reaches its maximum in the afternoon and then decreases back to zero again. The PV power exhibits similar correlated non-linear behavior which varies from day to day. Movements of cloud cover can cause sudden drop in the power generation.

This work supported in part by US Department of Energy (DOE) under grant DE-0E000060 and the US National Science Foundation (NSF) under grants #1312260, #1308192 and #1232070. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of Department of Energy and National Science Foundation.

Fig. 1 shows the variation of solar irradiance captured on June 5, 2014 by Real-Time Power and Intelligent System Laboratory (RTPIS Labs)'s weather monitoring system at Clemson University, South Carolina. A 200 MW PV plant was simulated on the real-time digital simulator (RTDS) using real-time weather profiles. Fig. 2 shows the variation of simulated PV power using solar irradiance in Fig. 1.

This stochastic behavior of the PV output is a challenge in predicting PV generation. For very short time intervals such as a minute, the PV output exhibits a huge non-linearity. Very short term PV power forecasts for prediction intervals ranging from few seconds to a minute plus are very important in making decisions and improving performamce of grid operations in electric utility control centers.

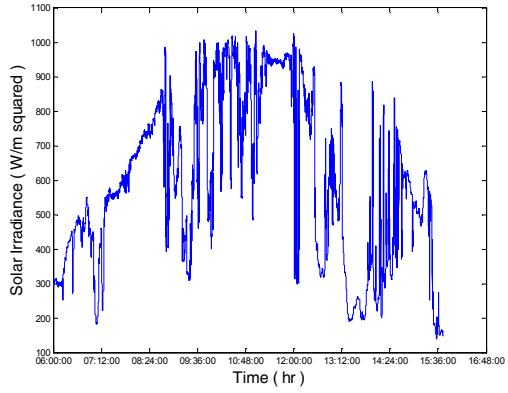

Fig. 1. Variation of solar irradiance on June 5, 2014.

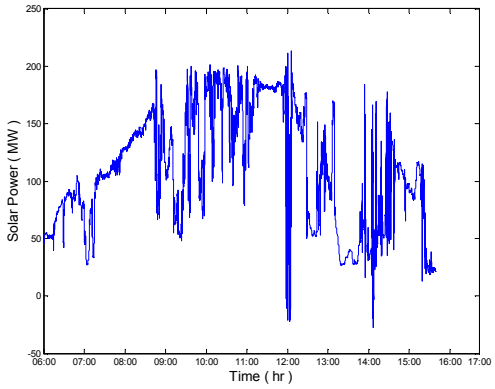

Fig. 2. Variation of RTDS simulated PV plant power output on June 5, 2014.

978-1-4799-4546-7/14/$31.00 ©2014 IEEE

There are numerous methodologies proposed for long term and short term forecasting of PV power. These methods include using mathematical equations [1], regression analysis [2] and linear time series models [3]. These methods are not optimal for PV power prediction, as output changes in a non-linear stochastic manner. Artificial neural networks (ANNs) have been used widely for nonlinear stochastic prediction such as power system loads [4] and wind speed [5]. Previous research in predictions uses ANNs including feedforward neural networks (FFNNs) and feedback neural networks [6]. In [6], a dynamic neural network is used to predict PV power output. Radial basis function (RBF) [7] and neural tree [8] are the other methods used to predict PV power. All these networks are difficult to train using complex input signals to provide sufficient prediction accuracy.

Echo state network (ESN) is a form of recurrent neural network (RNN) that was developed by Jaeger [9] in order to overcome difficulties in training. ESN has been used in predicting short to medium range prediction of solar irradiance [10]. In this paper an ESN is investigated to predict PV power output at multiple time instances ranging from a few seconds (5, 10, 15, 20, 25, 30) to a minute plus (60, 90 seconds) into the future. The performance of the ESN is compared with that of an extreme learning machine (ELM).

The remainder of this paper is organized as follows. ESN and ELM architectures are described in Section II. In Section III, application of ESN and ELM for output power prediction of a large PV plant is described. Typical results and discussions are provided in Section IV. Finally, concluding remarks and future work are given in Section V.

II. Reservoir Based Learning Networks

In this section, two types of reservoir based learning networks are described, namely, the echo state networks and the extreme learning machines.

A. Echo State Network (ESN)

Fig. 3 shows the basic network architecture of an echo state network. ESN has a fixed and large reservoir which contains randomly and sparsely connected neurons. The reservoir is driven by input signals ($U(n)$) and projects to output units ($Y(n)$). During learning, only the connections from the reservoir to these output units (readouts, W_{out}) are computed, leaving the rest of the network connections static with their randomly generated parameters (W_{in}, W and W_{fb}). Therefore ESN can be implemented without adjusting the weights between input-layer and the reservoir [9], [11], [13]. This makes the learning with ESN simpler and faster than with conventional RNN [12]. For the computation of readouts simple linear regression is used [9].

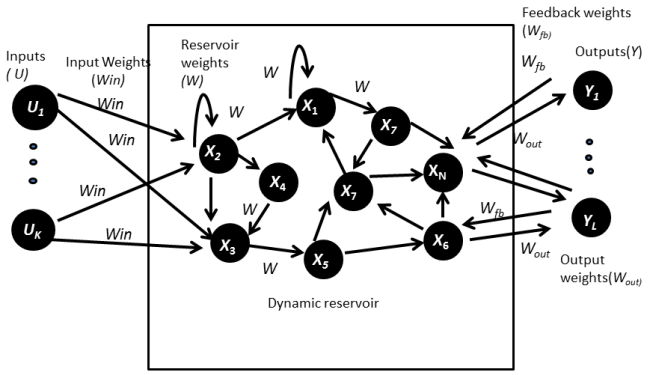

Fig. 3. An Architecture of an ESN.

Consider a discrete-time neural network with K input units, N internal network (reservoir) units and L output units.

Activations of input units at time step n are $U(n) = (U_1(n),...,U_K(n))$, of internal units are $X(n) = (X_1(n),...,X_N(n))$, of output units are $Y(n) = (Y_1(n),...,Y_L(n))$, and of teacher output units are $D(n) = (D_1(n),...,D_L(n))$ [11]. Dimensions of the weight matrices are as follows:

- Input weights matrix $W_{in} = [N \times K]$
- Reservoir weights matrix $W = [N \times N]$
- Feedback weights matrix $W_{fb} = [N \times L]$
- Readout weights matrix $W_{out} = [L \times (K + N)]$.

The activation of reservoir units for the n+1[th] time step is updated according to (1) [11].

$$X(n+1) = f(W_{in}U(n+1) + WX(n) + W_{fb}D(n)) \quad (1)$$

where f is activation function for reservoir neurons. Readout weights matrix is computed using (2).

$$(W_{out})^T = pinv(M) \times T \quad (2)$$

where M is the matrix of concatenated input/reservoir states $(U(n), X(n))$ as a row and $pinv(M)$ is the Moore-Penrose pseudoinverse of M. T is the sigmoid-inverted teacher output $tanh^{-1}(D(n))$ as a row. Finally, the output is computed according to (3).

$$Y(n+1) = f_{out}(W_{out}(U(n+1), X(n+1)) \quad (3)$$

where f_{out} is activation function for output neurons.

The reservoir should have important properties for the learning performance and is created to be rich in dynamics [14]. The most important property is echo state property, which states that the reservoir will asymptotically wash out any information from initial conditions (settling time) [12].

A necessary and a sufficient algebraic condition on the reservoir weight matrix are known, which ensure the echo state property. Generally majority of the reservoir weights are set to zero, providing sparse connectivity. A small percentage (often as low as 5%) of the reservoir weights are assigned to non-zero values, either in a random fashion or through probabilistic selection from few preselected values [14]. Additionally, the three most important global control parameters (reservoir network size, spectral radius of reservoir weight matrix, scaling of inputs) must be considered to improve the performance of the ESN [9], [14], [15]. The fixed values for input weights are mostly non-zero, providing dense connectivity. Also if the output feedback weights are used, mostly non-zero values are used.

B. *Extreme Learning Machine (ELM)*

ELM is a single hidden layer feedforward neural network (SLFN) which randomly chooses hidden nodes and analytically determines the output weights. This nature of ELM answer the most identified problem in feedforward networks (FFN) that is slow learning speed due to slow gradient based learning algorithms. The most significant characteristic of the ELM is that it provides good generalization performance at extremely fast learning speed [16]. Fig. 4 shows the basic network architecture of an ELM.

ELM can be described in the following way. Consider n arbitrary distinct samples $(U(i), Y(i))$ where $U(i) = [U_1(i),...,U_K(i)]^T \in R^K$ and $Y(i) = [Y_1(i),...,Y_L(i)]^T \in R^L$,activation function g and N number of hidden nodes. The steps in creating an ELM are:

- Randomly assign input weights W_{in} and bias b_i,.where $W_{in} = [W_{in1}, ..., W_{inN}]^T$
- Calculate hidden layer matrix $H=[n \times N]$

$$H(w_{in1},...,w_{inN},b_1,...,b_N,U(1),...,U(n)) = \begin{pmatrix} g(w_{in1}U(1)+b_1) & \cdots & g(w_{inN}U(1)+b_N) \\ \vdots & \ddots & \vdots \\ g(w_{in1}U(n)+b_1) & \cdots & g(w_{inN}U(n)+b_N) \end{pmatrix} \quad (4)$$

- Calculate the output weight W_{out} as

$$W_{out} = pinv(H) \times T \quad (5)$$

$$W_{out} = [N \times L] = [W_{out1}^T;...;W_{outN}^T] \quad (6)$$

$$T = [N \times L] = [Y(1)^T;...;Y(N)^T] \quad (7)$$

where *pinv(H)* stands for Moore Penrose pseudoinverse of H. The number of hidden nodes N must be estimated correctly to obtain better results [17].

- Finally, calculate the output according to (8)

$$Y = H \times W_{out} \quad (8)$$

III. PV Power Output Prediction

This section describes the experimental setup used in this study to generate the inputs and targets for the ESN and ELM learning.

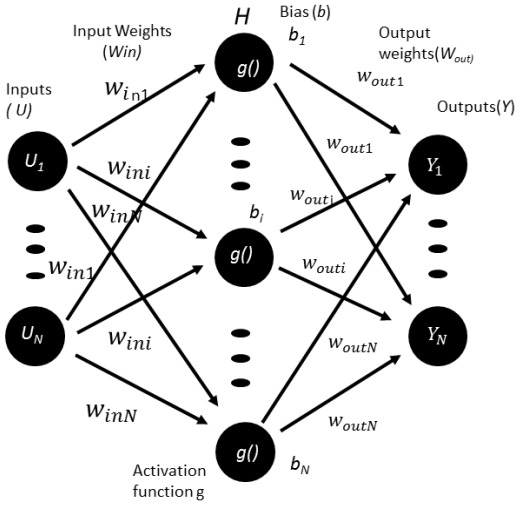

Fig. 4. An Architecture of an ELM.

Furthermore, the ESN and ELM development for PV power output prediction are explained.

A. *PV Plant System and RTPIS Weather Monitoring System*

A large 200MW PV plant is simulated on the-RTDS for the power system using four 50 MW PV plants as shown in Fig. 5. The RTPIS lab's weather monitoring system captures several weather parameters on a second basis including temperature, $T(t)$, and solar irradiance, $I(t)$. These weather parameters are passed onto the RTDS to implement a real-time simulation of the PV plant power outputs. Corresponding PV power outputs $P_{PV}(t)$, are captured for the respective temperature and irradiance measurements for the ESN and ELM developments.

B. *ESN for PV Power Prediction*

Temperature, solar irradiance and PV power output at current time are used as inputs. In this study, the PV power prediction is very short term, in other words in the order of few seconds to a minute plus (5s, 10s, 15s, 20s, 25s, 30s, 60s, and 90s). A single dynamic reservoir is exploited to provide eight predictions at different time instances with the same input and reservoir. Reservoir matrix is updated as in (1). The ESN is tested with feedbacks and without feedbacks, and observed that the ESN performs well without feedbacks. Readout weights are updated using (9) as per (2).

$$(W_{out}(t))^T = pinv(M(t)) \times T(t) \quad (9)$$

where t represents time steps 5s, 10s, 15s, 20s, 25s, 30s, 60s and 90s. $W_{out}(t)$ is the readout weight at time step t, $T(t)$ is the sigmoid inverted teacher output at time step t, $M(t)$ is the matrix M at time step t and $pinv(M(t))$ is the Moore Penrose pseudoinverse of $M(t)$.

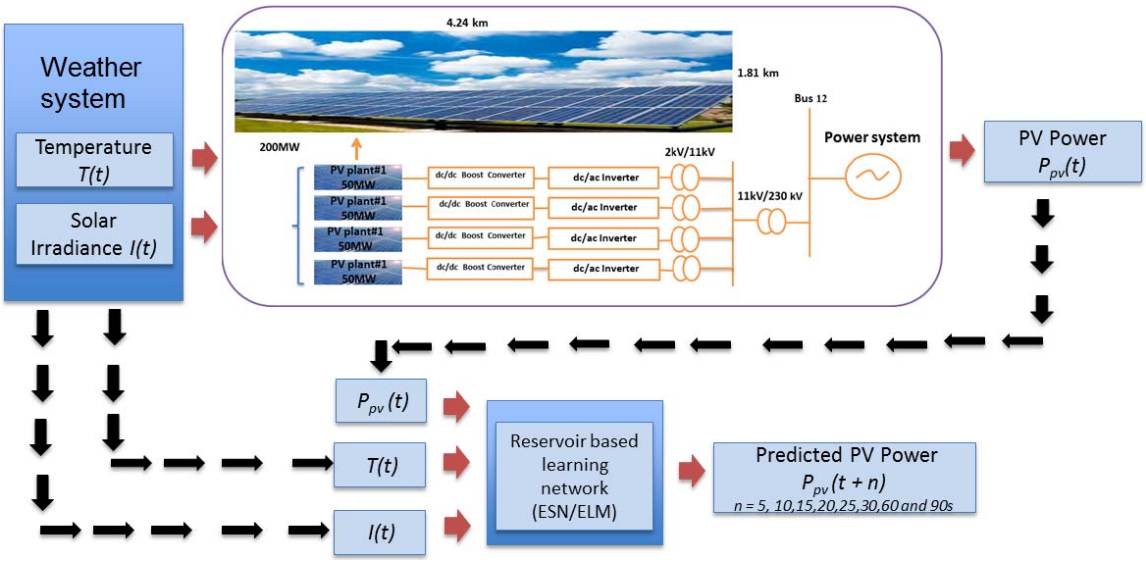

Fig. 5. Schematic diagram of a real-time similation of a large PV plant consisting of four 50 MW PV plants with actual weather (Clemson, SC) on RTDS.

A hyperbolic tangent activation function is used for both reservoir and output neurons. The created ESN is tested with different global parameter values [18]; the most performing ESN has 100 reservoir neurons. Reservoir weights are populated in a manner that guaranteed the echo state property. It is populated by setting 1% of the matrix in to a randomly generated value between [-1, 1] and spectral radius is set to 0.8 [14]. Input weights are populated based on the number of inputs and reservoir neurons. Feedback weights (W_{fb}) are populated using randomly selected values from a uniform distribution ranging [0, 1] and the settling time of the reservoir is set to be within the first 100 samples.

C. ELM for PV Power Prediction

ELM is developed so that the hidden layer matrix is capable of calculating eight different outputs using eight different output weights. Same input parameters are used as in the ESN. Calculation of output weights is carried out using (10) as per (5).

$$W_{out}(t) = pinv(H) \times T(t) \tag{10}$$

where t represents time steps 5s, 10s, 15s, 20s, 25s, 30s, 60s and 90s. $W_{out}(t)$ and $T(t)$ are the output weight and the teacher matrix (given in (7)), respectively, at time step t.

ELM is tested with different number of hidden layer neurons, the most performing ELM has 100 hidden neurons. Input weights are populated based on the number of inputs and hidden layer neurons.

IV. Results and Discussion

A. Data Representation and Preprocessing

For learning and testing phases, one day of daytime data is used. For the learning phase, a data set between 6.00 AM and 6.00 PM that was captured on June 21 2014 was used. For the testing phase, a dataset between 6.00 AM and 6.00.PM that was captured on June 22 2014 was used. The sampling time was a second in these data sets. Thus, the data sets had a total of 43200 data points.

1) Normalization

All the learning and testing data points are normalized to the range of [0.01,1.01] by using the feature scaling function which is also called unity based normalization [19].

$$X_{norm} = a + \frac{(X - X_{min})(b - a)}{(X_{max} - X_{min})} \tag{11}$$

where X_{norm} is the normalized data point, X is the original data point, X_{min}, X_{max} are minimum and maximum values of the dataset respectively. [a, b] represents the range of normalization, in this study this is [0.01, 1.01].

2) Performance evaluating measures

The absolute percentage error (APE), mean absolute percentage error (MAPE) and correlation coefficient (ρ) are used to measure the performance of the implemented network models.

$$APE_i = \left| \frac{D(i) - Y(i)}{D(i)} \right| \times 100\% \tag{12}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} APE_i \tag{13}$$

$$\rho = \frac{\sum_{i=1}^{n} (Y(i) - \bar{Y})(D(i) - \bar{D})}{\sqrt{\sum_{i=1}^{n} (Y(i) - \bar{Y})^2} \sqrt{\sum_{i=1}^{n} (D(i) - \bar{D})^2}} \tag{14}$$

where n is the number of data points, $D(i)$ and $Y(i)$ are i^{th} target value and i^{th} predicted value respectively. APE_i is the APE for i^{th} sample point. $\overline{Y}$ and $\overline{D}$ represents the mean values of the target and predicted value respectively.

B. Results

Figs. 6 and 7 show the learning and testing data distribution for (t+90)s prediction time step obtained by ESN and ELM respectively. Following sections describe the results obtained from the performance measures described above.

1) Absolute Percentage Error(APE)

Figs. 8 and 9 compares the APEs obtained for time step 90s from ESN and ELM respectively. Y axis of the graph is capped at 20% APE to visualize the variation clearly. APEs are calculated for each data point in the data set. Then the APE values are sorted in ascending order and 95% of the data points are selected to calculate mean absolute percentage error (MAPE). Remaining 5% of data set is neglected by assuming those points are outliers. Maximum APE values of 95% of data points for eight time steps are given in Table I.

According to the Table I, maximum APEs for ELM are better than that of ESN for time steps t+30, t+60 and t+90 for both learning and testing phases. The difference between maximum APE of ESN and ELM is large for training phase compared to testing phase for these time steps. It is harder to train ESN to predict large time step values due to its complex, highly non-linear reservoir. During the testing phase, the values are approximately equal. ESN reservoir retains better performance during testing. It can be observed from Figs. 10, 11 and 12 that maximum APEs are obtained with ESN. However, about 90% of ESN APEs are less then APEs with ELM. This implies that ESN performs well than ELM for most time steps. Additionally, further research is required to improve the training of complex ESN networks.

2) Mean Absolute Percentage Error(MAPE)

Both the ESN and ELM models are executed for 20 trials and the mean of APEs for 95% of the data set (5% of outliers are removed) are obtained for eight time step predictions. In Table II, the MAPEs obtained for ESN and ELM for all eight time step predictions are compared.

3) Correlation Coefficient

Table III shows the correlation coefficients for target versus predictions obtained for ESN and ELM for all the time steps (100% of data set). Figs. 13 and 14 show the correlations obtained for the time step t+30s prediction during the learning phase by ESN and ELM respectively. Figs. 15 and 16 show the correlation of the results obtained for the same time step when the 5% of outlier data is neglected.

TABLE I. MAXIMUM APE FOR 95% OF DATA

Prediction at time *t* for time instant	Learning APE		Testing APE	
	ESN (%)	ELM (%)	ESN (%)	ELM (%)
t+5	4.0510	28.5779	7.1648	28.6682
t+10	6.5555	22.5250	12.2297	22.5588
t+15	6.6136	19.6154	12.8436	19.9516
t+20	8.4515	17.1191	15.0058	17.9940
t+25	10.4764	12.7942	17.4259	17.3319
t+30	12.9866	11.1254	19.7346	19.2508
t+60	25.4340	16.3513	30.8825	29.7368
t+90	32.0149	26.6144	37.7495	37.6847

TABLE II. MAPE FOR LEARNING AND TESTING PHASES

Prediction at time *t* for time instant	Learning MAPE		Testing MAPE	
	ESN (%)	ELM (%)	ESN (%)	ELM (%)
t+5	**0.7221**	3.8587	**1.1954**	4.4389
t+10	**1.3590**	3.7714	**2.3811**	4.5701
t+15	**1.5381**	3.6587	**2.5328**	4.6934
t+20	**1.8416**	3.5934	**3.0215**	4.7822
t+25	**2.2709**	3.3998	**3.6592**	4.5902
t+30	**2.5613**	3.1314	**3.9442**	4.3882
t+60	**4.2741**	4.4787	**6.0993**	6.3959
t+90	**5.4328**	6.2039	**7.6080**	8.6509

TABLE III. CORRELATION COEFFICIENT FOR LEARNING AND TESTING PHASES

Prediction at time *t* for time instant	Learning Correlation Coefficient		Testing Correlation Coefficient	
	ESN	ELM	ESN	ELM
t+5	**0.9874**	0.9846	**0.9907**	0.9883
t+10	**0.9781**	0.9765	**0.9843**	0.9827
t+15	**0.9662**	0.9636	**0.9777**	0.9764
t+20	**0.9566**	0.9534	**0.9727**	0.9714
t+25	**0.9484**	0.9448	**0.9691**	0.9677
t+30	**0.9430**	0.9381	**0.9650**	0.9630
t+60	**0.9245**	0.9152	**0.9451**	0.9429
t+90	**0.9153**	0.9040	**0.9277**	0.9242

V. CONCLUSION

Two learning networks for predicting PV power outputs for different time steps ahead are presented in this paper. The performance of the echo state network was compared with that of the extreme learning machine. The performance measures have provided several insights to ESN and ELM based on PV power predictions. Overall, the ESN has better performance than the ELM in terms of prediction accuracy measures - APE, MAPE and correlation coefficient. This is due to the existence of a rich dynamic reservoir in ESN. Future work includes expanding the input parameters to have cloud cover and relative humidity measures. For example, cloud cover information can be collected from distributed sensors placed in proximity to a PV plant. Furthermore, closed-loop control of a power system with large PV plants and tie-line power flow control with predicted power output of PV plant can improve the frequency regulation in a multi-area power system.

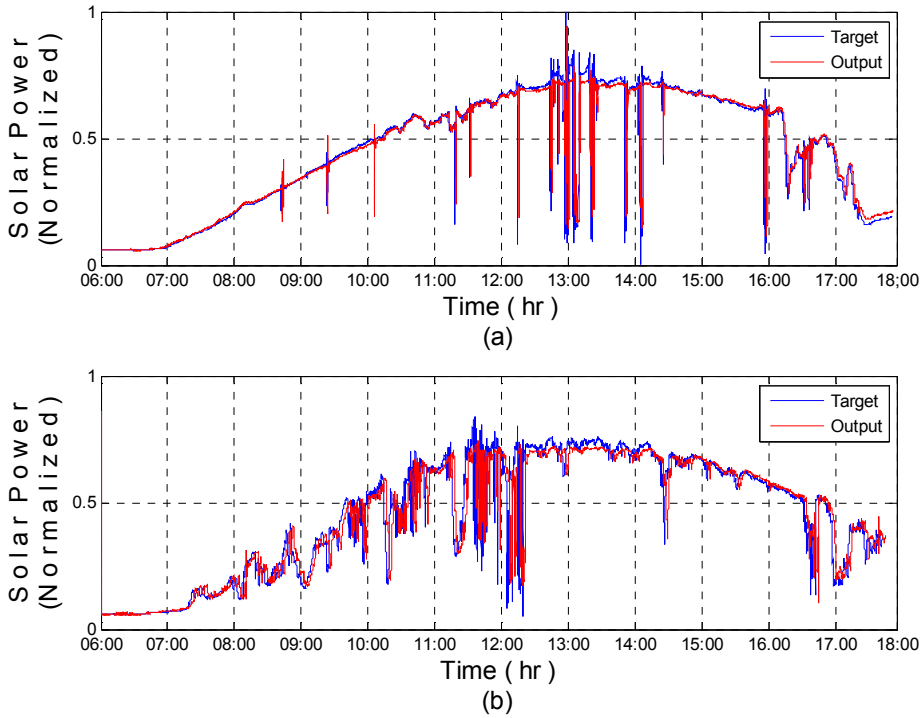

Fig. 6. (a) Learning target and output data distribution for prediction time step 90s obtained by ESN and (b) Testing target and output data distribution for prediction time step 90s obtained by ESN.

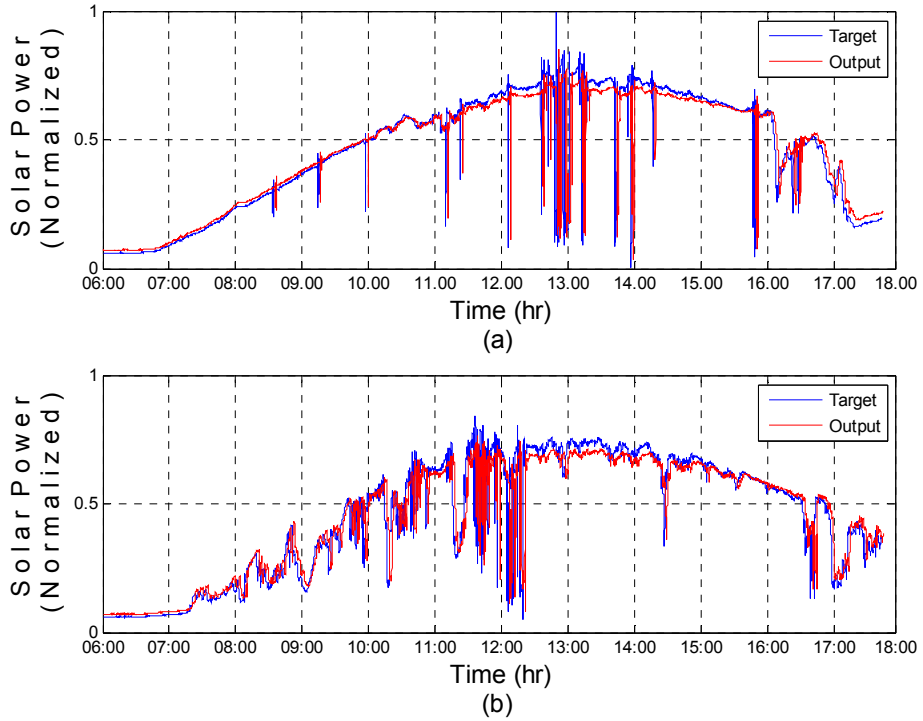

Fig. 7. a) Learning target and output data distribution for prediction time step 90s obtained by ELM and (b) Testing target and output data distribution for prediction time step 90s obtained by ELM.

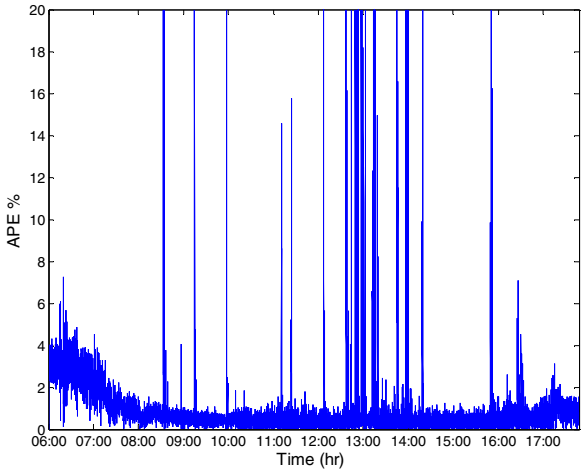

Fig. 8. Learning APE for (t+90)s prediction time step obtained using ESN.

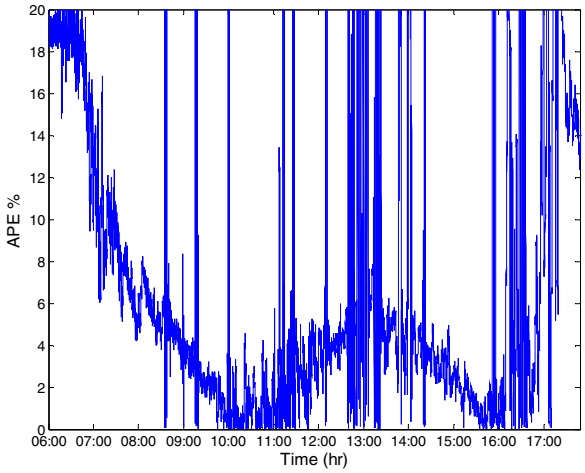

Fig. 9. Learning APE for (t+90)s prediction time step obtained using ELM.

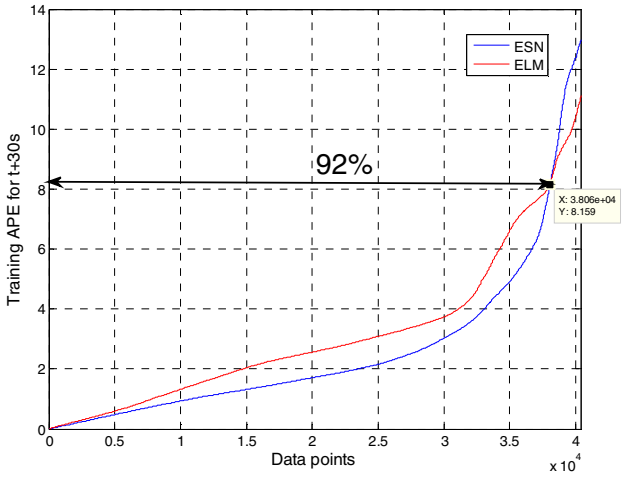

Fig. 10. Sorted traning APEs(from 95% data set) for (t+30)s prediction timestep

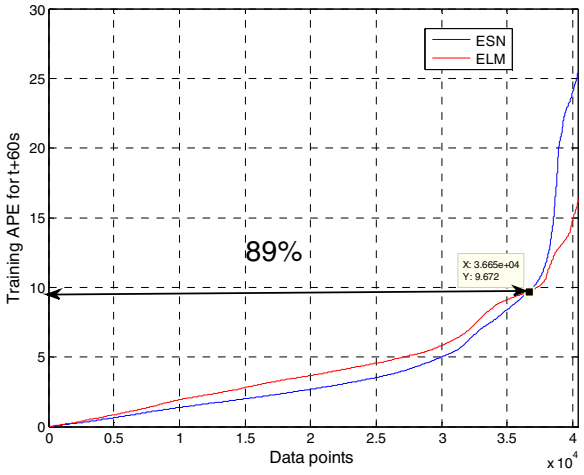

Fig. 11. Sorted traning APEs(from 95% data set) for (t+60)s prediction timestep

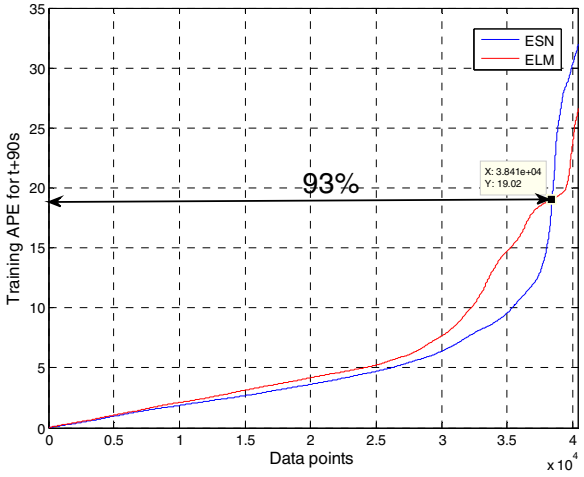

Fig. 12. Sorted traning APEs(from 95% data set) for (t+90)s prediction timestep

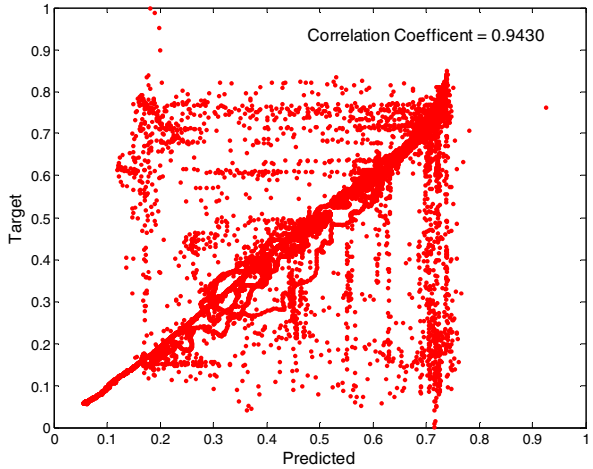

Fig. 13. Correlation coefficent (for 100% of data set) for (t+30)s prediction time step during the ESN learning phase

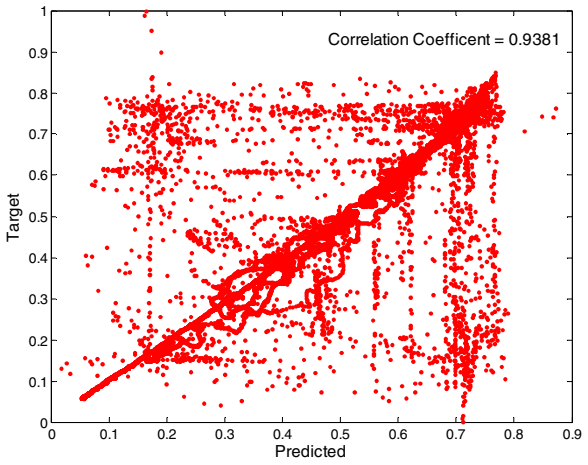

Fig. 14. Correlation coefficent (for 100% of data set) for (t+30)s prediction time step during the ELM learning phase .

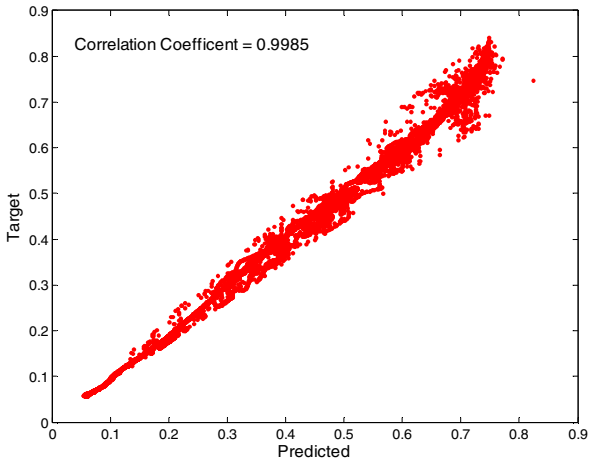

Fig. 15. Correlation coefficent (for 95% of data set) for (t+30)s prediction time step during the ESN learning phase.

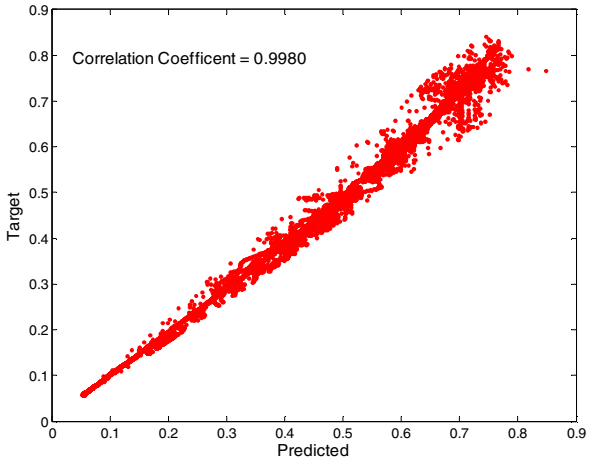

Fig. 16. Correlation coefficent (for 95% of data set) for (t+30)s prediction time step during the ELM learning phase.

REFERENCES

[1] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States," Solar Energy, vol. 855, p. 967–977, 2011.

[2] C. Voyant, C. Paolia, M. Musellia and M.-L. Niveta, "Multi-horizon solar radiation forecasting for Mediterranean locations using time series models," Renewable and Sustainable Energy Reviews, vol. 28, pp. 44-52, 2013.

[3] P. Bachera, H. Madsena and H. A. Nielsenb, "Online short-term solar power forecasting," Solar Energy, vol. 83, no. 10, p. 1772–1783, 2009.

[4] B. Kermanshahi and H. Iwamiya, "Up to year 2020 load forecasting using neural nets," International Journal of Electrical Power & Energy Systems, vol. 24, no. 9, p. 789–797, 2002.

[5] Q. Caoa, B. T. Ewinga and M. A. Thompsona, "Forecasting wind speed with recurrent neural networks," European Journal of Operational Research, vol. 221, no. 1, p. 148–154, 2012.

[6] N. Al-Messabi, L. Yun, I. El-Amin and C. Goh, "Forecasting of photovoltaic power yield using dynamic neural networks," in Neural Networks (IJCNN), Brisbane, QLD, 2012.

[7] J. Zeng and W. Qiao, "Short-Term Solar Power Prediction Using an RBF Neural Network," in IEEE Power and Energy Society General Meeting (PES), 2011.

[8] S. Basterrech, L. Prokop, T. Burianek and S. Misak, "Optimal design of neural tree for solar power prediction," in Electric Power Engineering (EPE), 15th International Scientific Conference, Brno-Bystrc, Czech Republic, 2014.

[9] H. Jaeger, "The “echo state” approach to analysing and training recurrent neural networks with an Erratum note," German National Research Center for Information Technology GMD Report 148, 2001.

[10] S. M. Ruffing and G. K. Venayagamoorthy, "Short to medium range time seres prediction of solar irradiance using an echo state network," in Intelligent System Applications to Power Systems (ISAP) 15th International Conference , Curitiba, 2009.

[11] H. Jaeger, "A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach," Fraunhofer Institute for Autonomous Intelligent Systems (AIS), 2013.

[12] H. Jaeger, "Short term memory in echo state networks," German National Research Institute for Computer Science, GMD-Report 152, 2002.

[13] H. Jaeger , "Scholarpedia Echo state network," 2007. [Online]. Available: http://www.scholarpedia.org/article/Echo_state_network. [Accessed 10 06 2014].

[14] G. K. Venayagamoorthy and S. Bashyal, "Effects of spectral radius and settling time in the performance of echo state networks," in Neural Networks, 2009.

[15] H. Jaeger, "Reservoir riddles: suggestions for echo state network research (extended abstract)," in International Joint Conference on Neural Networks (IJCNN), 2005.

[16] G. Huang, Q. Zhu and C. Seiw, "Extreme learning machine: theory and applications," in Neurocomputing 70, 2006, p 489-501

[17] S. Salcedo-Sanz, C. Casanova-Mateo, A. Pastor-Sanchez and M. Sanchez-Giron, "Daily global solar radiation prediction based on hybrid coral reefs optimization – extreme learning machine approach," in Solar Energy 105, 2014.

[18] M. Lukoševičius, "A practical guide to applying echo state networks," in Neural Networks Tricks of the Trade, 2nd edition, Springer, 2012, pp. 659-686.

[19] "Wikipedia, Normalization (statistics)," 2012. [Online]. Available: http://en.wikipedia.org/wiki/Normalization_(statistics). [Accessed 12. 06. 2014].