Detecting Wind Power Ramp with Random Vector Functional Link (RVFL) Network

Conference Paper · December 2015

DOI: 10.1109/SSCI.2015.105

4 authors, including:

Xueheng Qiu
Nanyang Technological University
16 PUBLICATIONS **696** CITATIONS

SEE PROFILE

Ponnuthurai N. Suganthan
Nanyang Technological University
522 PUBLICATIONS **43,106** CITATIONS

SEE PROFILE

Gehan Amaratunga
University of Cambridge
778 PUBLICATIONS **29,888** CITATIONS

SEE PROFILE

Detecting Wind Power Ramp with Random Vector Functional Link (RVFL) Network

Ye Ren, Xueheng Qiu, and P. N. Suganthan
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore
{*re0003ye, qiux0004, epnsugan*}@ntu.edu.sg

Narasimalu Srikanth
Energy Research Institute @ NTU (ERI@N)
Nanyang Technological University, Singapore
nsrikanth@ntu.edu.sg

Gehan Amaratunga
Department of Engineering
University of Cambridge, UK
gaja1@hermes.cam.ac.uk

***Abstract*—Due to the intermittent nature of the wind, the wind speed is fluctuating. Fluctuating wind speed cause even more fluctuation in wind power generation. The sudden changes of the wind power injected into the power grid within a short time frame is known as power ramp, which can be harmful to the grid. This paper presents algorithms to detect the wind power ramps in a certain forecasting horizon. The importance and challenges of wind power ramp detection are addressed. Several different Wind power ramps are defined in this paper. A random vector functional link (RVFL) network is employed to predict the future occurrence of wind power ramp. The forecasting methods are evaluated with a real world wind power data set. The RVFL network has comparable performance as the benchmark methods: random forests (RF) and support vector machine (SVM) but it has better performance than the artificial neural network (ANN). The computation time of training and testing is also in favor of the RVFL network.**

I. Introduction

Due to the pollution caused by fossil fuel power generation and the exhausting fossil fuel power sources, renewable energy sources such as wind and solar draw more and more attention of the industry and government. Wind energy has a vast development in the past few years and will continue to expand in the future. However, due to the intermittent nature of the wind, the power generated by wind turbines is fluctuating. The large change of output power in a short time interval is known as ramps [1], [2]. The fluctuations in the wind power is usually compensated by the conventional fossil power generator or battery storage systems. If the ramp exceeds beyond the buffer of the power reservoir, the compensation will be less effective.

Accurate wind power ramp detection is beneficial to the power system planning and scheduling from sudden power change [3], [4] and to protect the power transmission and generation system from a sudden rise and drop in power supply [5].

By further looking at the nature of the ramp and it can be categorized into two types: up ramp and down ramp. The up ramp occurs when the power generated significantly increases in a short period and the down ramp occurs when the power generated significantly decreases in a short period [6]. However, since complete power system vary in size and customer, there is no industrial standard to clearly define a ramp quantitatively [7]. Therefore we are going to provide some power ramp definitions that are closely related to the real world wind power dataset used in this paper.

978-1-4799-7560-0/15/$31

Various research focusing on the identification of wind power ramps are given in [8]–[10]. These research papers serve as a survey to different definitions of wind power ramps. Methodology to classify wind power ramps are also presented in the literature such as [3], [6], [7]. In [3], an artificial neural networks (ANN) was employed to classify the wind power ramp. The classification is based on the stochastic process and the output is not rigid class labels but the probability. Another popular classification method called support vector machine (SVM) was also used to forecast the wind power ramp in [2]. By convert categorical valued wind power ramps into continuous valued wind power ramp rates, some conventional regression methods can be applied [7] such as support vector machine (SVM), random forests (RF) and ANN. Some wind speed and wind power forecasting methods [11], [12] can be imported to forecast wind power ramp rate as well.

Random vector functional link (RVFL) network [13]–[15] is a variation of ANN that has (i) direct connections from the input layer to the output layer and (ii) the weights of the interconnections between the input and the hidden layer are randomly assigned and (iii) the weights of the interconnections from the input and hidden layers to the output layer are optimized by a least square method. With the three novelties, it has faster training time than the conventional ANN but without significant trade-off in accuracy.

In this paper, the RVFL network is used to detect the occurrence of wind power ramps in the next 6 and 12 hours based on binary class classification. Another objective is to propose a procedure to process the imbalanced wind power ramp data including balancing and noise detection and smoothing.

The remaining of the paper is organized as follows: Section II defines the wind power ramp that are used in this paper; Section III introduces imbalanced data classification; Section IV details the RVFL network; Section V shows the experimental results and discusses the performances of the RVFL network based on the results and finally Section VI concludes the paper and give recommendations for future work.

II. Wind Power Ramp

A wind power ramp is the change of power, either increasing or decreasing in a specific time. We call the increasing power ramp up ramp and the decreasing power ramp down ramp. Wind power ramp is common in wind power generation because the wind speed is fluctuating. In addition, wind power is not a linear function with respect to wind speed because it has cut in region, cut out region, cubic region and maximum-output region. A typical wind speed to wind power mapping is shown in Figure 1, and the data is taken from NREL west wind [16].

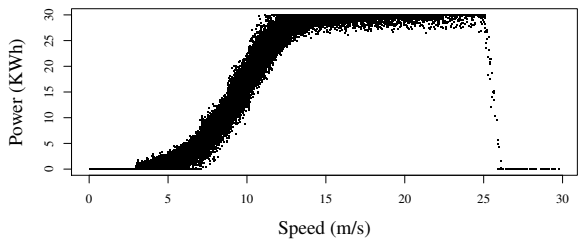

Fig. 1: Wind Speed to Wind Power Mapping.

The power output in the cut in region is zero because the wind is not powerful to overcome the internal friction in the wind turbine. The power output in the cut out region is zero because in order to protect the wind turbine from over spinning and over heating, the turbine is halt during strong wind. The maximum-output region outputs a constant power irrespective of the wind speed due to the maximum output capability of the generator in the turbine. The cubic region shows a cubic power relation between wind speed and wind power which follows an equation shown in (1) [17].

$$P = \frac{1}{2}\rho_a A_t C_p v^3 \tag{1}$$

where ρ_a is the air density, A_t is the area of the turbine when rotating, C_p is the efficiency, and v is the up-wind speed.

In the literature, there are numerous definitions of power ramps, in this chapter we introduce two most popular definitions and employ them in our experiments. The first definition is based on the local extrema within a certain time interval. A significant power ramp occurs when the ramp Γ_{ext} exceeds a certain threshold [9]:

$$\Gamma_{ext} = max(P(t,\ldots,t+\Delta t)) - min(P(t,\ldots,t+\Delta t)) > \Gamma_{val} \tag{2}$$

where $P(t)$ is the wind power generated at time t, Δt is the time interval, and Γ_{val} is the threshold to determine whether the ramp is significant or not.

Another definition is based on the two end points of a sliding window [9] and the wind power is denoted as Γ_{end}.

$$\Gamma_{end} = |P(t+\Delta t) - P(t)| > \Gamma_{val} \tag{3}$$

Notice that the above-mentioned two definitions do not take into account the ramp directions (up-ramp or down-ramp) and hence the decisions are binary (ramp or no ramp). The multiple class case (up-ramp, down-ramp or no-ramp) is not discussed in this chapter.

Based on these two definitions, the wind ramp forecasting can be converted to a binary classification problem.

III. Wind Power Time Series Pre-processing and Classification

Wind power time series (TS) is affected by meteorological factors and power system factors and thus it is highly fluctuating. Noise are created when (i) there is a strong wind gust in the environment, (ii) the wind turbine is shut down for maintenance, or (iii) there are sensor fault or noise in the data acquisition and recording system. In order to avoid wrongly identifying wind ramps from the TS due to noise, the noise embedded in the wind power TS should be removed. There are three types of noise: (i) outliers, (ii) missing values and (iii) white noise corresponding to the above-mentioned possible noise causes. We propose to detect them separately and smooth the outliers and missing values and discard the white noise.

A. Outlier

To detect the outlier, there is a rolling window with a fixed width to roll along the TS. For each window, there is a segment of TS $\mathbf{Y}_w = \{y_i, y_{i+1}, \ldots, y_{i+w}\}$, where w is the window width. For each $\mathbf{Y}_w$, the median absolute deviation (MAD) is calculated (equation (4)) and the data point which is $y_{MAD} \leq T_{th} \times \text{MAD}$ is considered as an outlier [18].

$$\text{MAD} = \text{median}_i\,(\ |Y_i - \text{median}_j(Y_j)|\ ) \tag{4}$$

B. Missing Data

The missing data is recorded as 'NA', 'NaN', '?', '-' or '-999' in the dataset depend on the data recording standard. These missing data can be easily identified manually.

C. White Noise

The white noise is usually embedded into the TS. The characteristic of white noise is that it has a constant power spectral density. Depending on the probability distribution, white noise can be Gaussian distributed. or uniformly distributed.

Empirical mode decomposition (EMD) based method can be applied to detect the white noise [19], [20]. A TS can be decomposed into intrinsic mode functions (IMFs) and a residue. The characteristics of the IMFs are: (i) zero mean, (ii) the number of zero-crossings and the extrema differ at most by one and (iii) the IMFs are (nearly) orthogonal to each other [21].

Viewing from the frequency domain, the energy density of a single IMF is:

$$E_n = \frac{1}{N}\sum_{t=1}^{N} c_n(t)^2 \tag{5}$$

where N is the length of the TS, E_n is the nth IMF's energy density and $c_n(t)$ is the nth IMF.

The average period of an IMF $\bar{T}_n$ is calculated based on the Fourier spectrum weighted mean period [20].

After obtaining E_n and $\bar{T}_n$, the particular IMF can be checked because if it is a white noise, the relationship between the two variables are [19]:

$$\ln E_n + \ln \bar{T}_n = 0 \tag{6}$$

The derivation of the confidence intervals can be found in [19], [20] and the equation of the confidence interval is:

$$\ln E_n = -\ln \bar{T}_n \pm k\sqrt{\frac{2}{N}\exp\frac{ln\bar{T}_n}{2}} \tag{7}$$

where N is the length of the TS and k is the percentiles of a standard normal distribution,

However, if the average period $\bar{T}_n$ is calculated by counting the number of zero-crossings (or extrema), there is an empirical correction [19], [20] which is:

$$\ln E_n + 0.934 \ln \bar{T}_n = 0.12 \tag{8}$$

The IMFs whose energy and average period obey the equation is considered as a white noise and the particular IMFs are removed from re-construction back to the TS.

D. Imbalanced Data Classification

Classification is to predict a decision based on a set of features, and the decision can be either binary class or multiple class. For wind power ramp forecasting based on the definitions, the classification task is binary.

However, for the processed wind power ramp data, the classes are imbalanced, i.e. there are much more 'no ramp' cases (majority class) than 'ramp' cases (minority class), which misleads the training process because the accuracy is usually even higher when all predicted decisions are under the majority class, which is obviously a wrong conclusion. In order to overcome this problem, the majority class in the training data is under sampled to match the number of minority class, or the minority class in the training data is over sampled to match the number of majority class [22]. In addition, the criteria to measure the performance of the classifier are appropriately selected for imbalanced data as well.

E. Cross Validation

Cross validation is a common training method for supervised learning such as classification and regression. Without cross validation, the trained classification/regression model tends to fit the training data very well but has poor generalization performance on unseen data (testing data), which is called over-fitting. Cross validation is to provide a validation data set that will evaluate the performance of the trained model before applying to the testing data. Cross validation is usually executed k times to average out the uncertainties and this is called k-fold cross validation.

For classification, the k-fold cross validation randomly subsamples the training data set into k subsets and repeatedly uses the $k-1$ subsets for training and the remaining subset for validation. This procedure is usually applied over several parameter sets to choose the best parameters.

F. Performance Measures

The binary class classification model is evaluated based on a contingency table (or confusion matrix). The contingency table is a 2×2 matrix that reflects the relationship between the target and the predicted values. Derived from the contingency table are the several performance metrics as shown below.

TABLE I: Contingency Table to Evaluate the Performance of the Binary Class Classification

	$\hat{x}=+1$	$\hat{x}=-1$
$x=+1$	TP	FN
$x=-1$	FP	TN

Derived from the contingency table are there several performance metrics such as:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \tag{9}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{11}$$

$$\text{F Score} = \frac{2TP}{2TP+FP+FN} \tag{12}$$

where C is the class of the target data, $\hat{C}$ is the class of the predicted data, TP, TN, FP and FN stand for true positive, true negative, false positive and false negative, respectively.

However, for imbalanced data, Accuracy is not a good metric to measure the performance and the preferred metrics are Precision, Recall and F score [6], [10]. Therefore, they are used in this paper to evaluate the wind power ramp classification methods.

IV. Random Vector Functional Link Neural Network

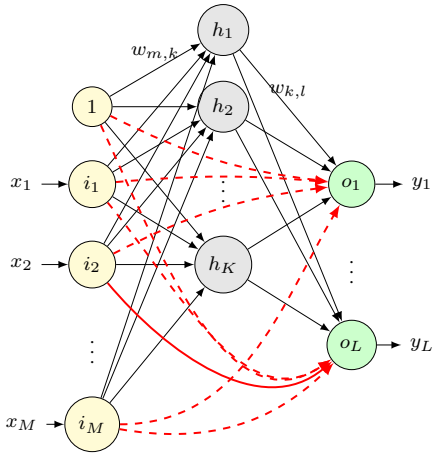

Fig. 2: Schematic Diagram of an RVFL Network, the dashed arrows show the direct connections between the input neurons and the output neurons.

ANN is a commonly used method for classification. The basic building block of an ANN is perceptron. A perceptron

has several input connections and an output connection and applies a nonlinear activation function to the inputs to calculate the output. The function of a perceptron is shown:

$$p_k = f(\sum_{m=1}^{M} w_{m,k} i_m + w_{m+1,k}) \tag{13}$$

where p_k is the kth perceptron's output value, $w_{m,k}$ is the weight of the connection between the mth input and the kth perceptron, M is the total number of inputs, $w_{m+1,k}$ is an input bias and $f(\cdot)$ is a non-linear activation function.

The commonly used activation function in the perceptron is sigmoid. There are several types of sigmoid functions in the literature and in this paper, we use logistic sigmoid ($logsig$) as shown:

$$\text{logsig}(x) = \frac{1}{1+e^{-x}} \tag{14}$$

There are a various types of ANN structures in the literature and among them, the most widely used is single-hidden layer feed-forward neural network (SLFN), and it is sometimes referred as one hidden layer multiple layer perceptron (MLP). A schematic diagram of an SLFN is shown in Figure 2.

As shown, there are three layers in the SLFN: input layer, hidden layer and output layer. The perceptrons in the hidden layer has the function as in equation (13). The perceptrons in the input layer is to feed the inputs to the hidden layer without any calculation. The perceptrons in the output layer is to calculate the weighted sum of the results of the hidden layer perceptrons as shown:

$$o_l = \sum_{k=1}^{K} w_{k,l} h_k + w_{k+1,l} \tag{15}$$

where o_l is the lth output perceptron's value, $w_{k,l}$ is the weight of the connection between the kth hidden layer perceptron and the lth output perceptron, K is the number of hidden layer perceptrons and $w_{k+1,l}$ is hidden layer bias.

To train an SLFN, the initial step is to assign random values to the weights $w_{m,k}$ and $w_{k,l}$, then apply certain method such as back-propagation (BP) [23] to tune the weights so that the outputs of the SLFN matches the target values with minimal error. However, BP-based methods are usually very slow in computation and the gradient descent is usually trapped in a local minimum instead of global minimum, thereby resulting in sub-optimal weights for the SLFN.

In order to fast train the SLFN, an SLFN with random weights was reported in [24]. Later in [13]–[15] an SLFN with direct input-output connections (functional link) was reported. The reported RVFL network does not employ BP to train the network but use random weights with least square estimation to obtain the optimal weights. The schematic diagram of an RVFL network is shown in Figure 2. Notice that the difference between RVFL network and SLFN is shown as the dashed connections from the input layer to the output layer.

The procedure of training an RVFL network is as follows [13], [15]:

1) Assign random values to input weights $w_{i,h}$. The random values are uniformly distributed in the interval $[0, 1]$.
2) Obtain the hidden perceptron outputs $\mathbf{A} = \text{logsig}(\mathbf{W}_{I,H} \cdot \mathbf{X})$, where $\mathbf{X}$ is the training data.
3) Apply least square estimation to calculate the output weights $w_{h,o}$ and the direct link weights $w_{i,o}$: $\mathbf{W}_O = (A^T A)^{-1} A^T \mathbf{Y}$, where $\mathbf{W}_O$ is the aggregation of output weights and the direct link weights, and $\mathbf{Y}$ is the training target.
4) Discard $\mathbf{W}_O$ from Step 3 and redo Steps 1 and 3 if there is rank deficiency in it.

The obtained $\mathbf{W}_O$ and $\mathbf{W}_{I,H}$ can then be applied to testing data to obtain the predicted values for the test data by:

$$\hat{Y}_s = \mathbf{W}_O \cdot \text{logsig}(\mathbf{W}_{I,H} \cdot \mathbf{X}_s) \tag{16}$$

where $\hat{Y}_s$ is the predicted testing values and $\mathbf{X}_s$ is the testing data.

Chen [14], [15] accelerated the training procedure by introducing sequential training. The RVFL network training procedures described in the previous paragraphs need to be executed whenever there is a change in the network structure, such as adding a hidden perceptron. Chen's sequential training enables the network to update the $\mathbf{W}_O$ and $\mathbf{W}_{I,H}$ based on the previously calculated data. The sequential training is also two-dimensional: one is to update the matrix when there is new training data presented to the network and the other is to update the matrix when there is an additional hidden perceptron added to the network. This sequential training made the RVFL network learning faster and having online learning capability.

A. Proposed Random Vector Functional Link Neural Network based Wind Power Ramp Forecasting Methods

We propose to use RVFL network [15] to classify the wind power ramp. There are two approaches and they are plotted in Fig. 3. The wind power TS first undergoes outlier removal, missing data removal and de-noising to be a smoothed TS $x(t)$. Firstly, the power ramp Γt is derived from the smoothed TS and the class labels are assigned to the it: '+1' for significant ramp and '-1' for no ramp. Then the power ramp series is partitioned into a training set and a testing set. In addition, the minority class in the training set is oversampled to Γ_t^b so that the amount of the two classes equal [22]. Next the RVFL network is trained with the training set and an optimal number of hidden layers determined by a k-fold CV to obtain a classification model m_c. Finally the trained model is applied to classify the testing data and the predicted power ramp $\hat{\Gamma}_s$ is obtained.

V. Results and Discussions

The wind power TS data is retrieved from ELIA wind power website [25]. The wind power TS spans from November 2014 to March 2015. The rated power of the wind farm is 712.9 MWh. The TS is partitioned into five monthly series named as $D1$ to $D5$, Each monthly series is sub-sampled from 15 min average to hourly average and scaled to $[0, 1]$ interval.

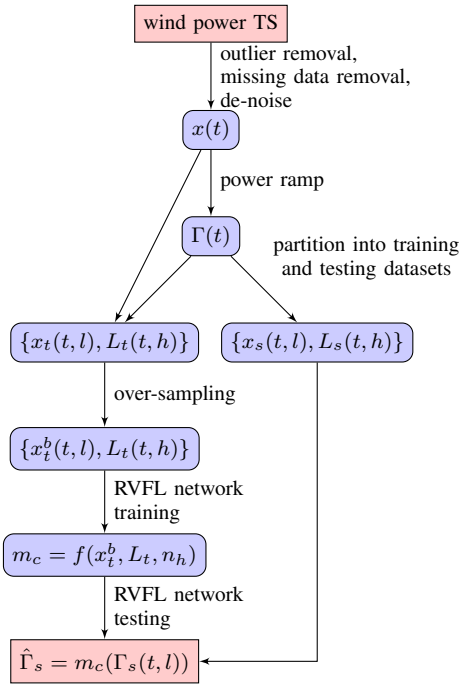

Fig. 3: Flowchart of RVFL Network on Wind Power Ramp Forecasting

The monthly TS are then partitioned into a first 50% training and remaining 50% testing structure.

The TS was smoothed by the MAD based outlier removal, missing data removal and moving average smoothing procedure. The TS was then decomposed by an EMD method and the energy and the average period of each IMF was calculated. If the two variable of the IMF followed equation (8), the IMF is discarded and the remaining IMFs and the residue is reconstructed back to the de-noised TS. The monthly TS were partitioned into a first 70% training and remaining 30% testing structure.

The $D1$ data is used to illustrate the EMD de-noising procedure. The energy v.s. average period is plotted in Fig. 4. We can see from that all the decomposed 8 IMFs neither fall onto the empirical white noise line (dashed straight line) and nor within the 1st-99th percentile confidence interval. Therefore, there is no IMFs considered as noise in this particular case.

Based on the wind ramp definition and the literature [8], we set $\Gamma_{val} = 25\%$ of the rated power (712.9 MWh) and $\Delta t = 4h$, we can find the start and end time points of the wind ramps based on the definition of Γ_{ext}. A fraction of wind power TS is plotted in Fig. 5 and the identified significant ramps are plotted in red segment lines by joining the start and end time points.

The classification task is to predict whether there exists a significant wind power ramp in the next 6 or 12 hours based on the previous 48 hours historical data. If there exists a wind power ramp, the class is '+1' (has ramp), otherwise the class is '-1' (no ramp). The training and target wind ramp decisions are pre-determined by the two wind ramp definitions: Γ_{ext}

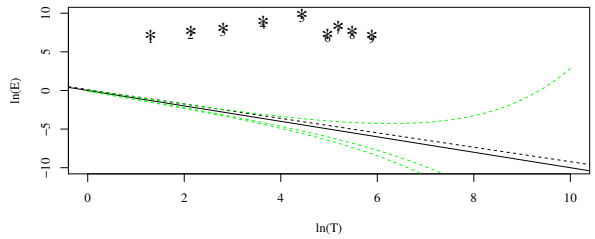

Fig. 4: The Energy v.s. Average Period Plot for D1 Data. The solid straight line is the theoretical line to determine white noise and the dashed straight line is the empirical line with average period calculated from counting zero-crossings and the two dashed curved lines are 5th and 95th percentile.

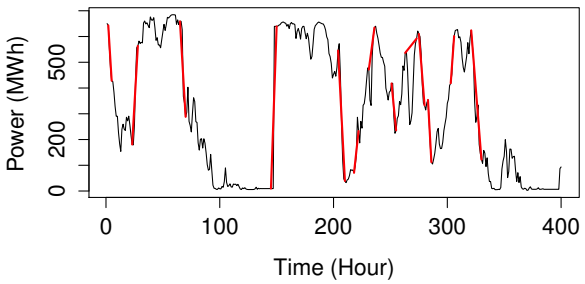

Fig. 5: A Fraction of Wind Power Generated in an ELIA wind farm, red segments denote power ramps.

and Γ_{end}. The summary of the datasets after partition, pre-processing and labeling is shown in Table II. It is shown that for the next 6 hour window forecasting, the datasets are imbalanced whereas for the next 12 hour window forecasting, the dataset is almost balanced. It is rational because longer time window means higher possibility to have wind power ramps falling inside. Therefore, over-sampling of the minority class in the training set is not necessary for 12 hour window forecasting.

A. Parameters Tuning of the Classification Methods

The RVFL network is evaluated with three other benchmark methods: ANN (single layer), SVM (RBF kernel) and RF. The parameters of RVFL, ANN and SVM are optimized by a 5-fold CV module and the RF has a build-in out-of-bag error module to select the optimal parameters. Instead of using accuracy as the parameter selection criterion, F score is applied.

The optimal parameters of ANN, SVM, and RVFL network after 5-fold CV and the parameters of RF are tabulated in Table III.

B. Classification Performance

The performance measures are tabulated in Table IV. The performance on 12 hour window forecasting is generally

TABLE II: Summary of the Datasets for Wind Ramp Classification

	6 Hour				12 Hour			
	Training		Testing		Training		Testing	
	+1	-1	+1	-1	+1	-1	+1	-1
Γ_{ext}								
D1	64	387	32	184	153	292	71	145
D2	332	136	55	168	252	210	111	112
D3	103	365	54	169	220	242	120	103
D4	77	340	58	144	152	259	75	127
D5	78	389	58	165	164	297	107	116
Γ_{end}								
D1	62	389	26	190	124	321	54	162
D2	130	228	45	178	237	225	85	138
D3	93	375	60	163	167	295	115	108
D4	82	335	78	124	150	261	62	140
D5	84	383	59	164	144	317	117	106

TABLE III: Optimal Parameters of the Classification Methods, 'mtry': number of variables randomly sampled as candidates at each split, 'ntree': number of trees to grow, n_h: number of hidden neurons, C: penalty factor of SVM, σ: RBF kernel's shape parameter.

		6 Hour					12 Hour				
		D1	D2	D3	D4	D5	D1	D2	D3	D4	D5
Γ_{ext}											
RF		mtry=16, ntree=500									
MLP	$n_h =$	96	42	96	54	54	48	54	96	54	30
SVR	$C =$	100	10000	100	100	1000	100	100	1000	100	100
	$\gamma =$	0.604	0.188	0.188	0.937	0.687	0.521	0.875	0.354	0.604	0.187
RVFL	$n_h =$	96	96	96	96	96	96	96	96	96	96
Γ_{end}											
RF		mtry=16, ntree=500									
MLP	$n_h =$	72	72	60	36	54	48	96	60	36	30
SVR	$C =$	100	1000	100	1000	100	100	100	100	100	100
	$\gamma =$	0.771	0.521	0.521	0.687	0.604	0.437	0.104	0.187	0.187	0.187
RVFL	$n_h =$	96	96	96	96	96	96	96	72	72	96

higher than that of the 6 hour future window forecasting which is due to the class balancing in the testing set. The performance measures of the RVFL is generally higher for both wind power ramp definitions on 12 hour window forecasting.

For 6 hour window forecasting. There are 'NaN' cases in F Score results due to the complete mis-classification of TP (also results a '0' in Precision and Recall). Although there is no 'NaN' or '0' from RVFL network, the performance is comparable among the four methods.

A Friedman rank sum test in Table V showed the statistical comparisons among the four classification methods. It is shown that for 6 hour window forecasting, the four methods are comparable whereas for 12 hour window forecasting, there are significant differences among them. Next a Nemenyi post-hoc test is applied to the 12 hour window forecasting and the p value is shown in Table VI. From the test statistics, we can see that RVFL network has comparable performances as ANN and RF and can outperform SVM for wind ramp defined by Γ_{ext}.

TABLE IV: Performance Measures of the Power Ramp Classification in the next 6 and 12 hours, 'NaN': not a number due to divide by zero.

		6 Hour				12 Hour			
	Dataset	ANN	RF	SVM	RVFL	ANN	RF	SVM	RVFL
F Score	Γ_{ext}								
	D1	0.031	NaN	0.07	**0.081**	0.17	0.162	0.149	**0.215**
	D2	0.226	0.135	0.146	**0.304**	0.458	**0.508**	0.365	0.478
	D3	0.189	0.123	0.204	**0.33**	0.324	0.512	0.43	**0.581**
	D4	**0.4**	0.304	0.222	0.392	0.579	0.631	0.543	**0.68**
	D5	0.314	**0.465**	0.253	0.296	0.362	**0.593**	0.331	0.585
Precision	D1	0.03	0	**0.08**	0.071	0.171	**0.286**	0.18	0.26
	D2	0.217	0.176	0.122	**0.233**	0.408	**0.473**	0.353	0.434
	D3	0.192	0.185	0.227	**0.234**	0.393	0.492	0.454	**0.531**
	D4	**0.351**	0.412	0.206	0.316	0.63	0.658	0.6	**0.691**
	D5	0.268	**0.714**	0.324	0.32	0.42	**0.683**	0.452	0.612
Recall	D1	0.031	0	0.062	**0.094**	0.169	0.113	0.127	**0.183**
	D2	0.236	0.109	0.182	**0.436**	0.523	**0.55**	0.378	0.532
	D3	0.185	0.093	0.185	**0.556**	0.275	0.533	0.408	**0.642**
	D4	0.466	0.241	0.241	**0.517**	0.535	0.606	0.496	**0.669**
	D5	**0.379**	0.345	0.207	0.276	0.318	0.523	0.262	**0.561**
F Score	Γ_{end}								
	D1	NaN	NaN	NaN	**0.061**	**0.122**	0.022	0.018	0.023
	D2	0.23	0.061	0.143	**0.265**	0.303	0.382	0.353	**0.483**
	D3	0.204	0.277	0.197	**0.329**	0.296	**0.54**	0.336	0.4
	D4	**0.453**	0.067	0.159	0.398	0.594	0.609	0.598	**0.636**
	D5	0.355	**0.509**	0.306	0.273	0.49	**0.578**	0.35	0.574
Precision	D1	0	0	0	**0.05**	**0.115**	0.027	0.017	0.029
	D2	**0.238**	0.095	0.111	0.198	0.312	0.333	0.303	**0.405**
	D3	0.182	**0.382**	0.194	0.272	0.341	**0.594**	0.352	0.419
	D4	**0.525**	0.25	0.183	0.357	0.679	**0.778**	0.713	0.755
	D5	0.338	**0.574**	0.385	0.247	0.51	**0.746**	0.519	0.583
Recall	D1	0	0	0	**0.077**	**0.13**	0.019	0.019	0.019
	D2	0.222	0.044	0.2	**0.4**	0.294	0.447	0.424	**0.6**
	D3	0.233	0.217	0.2	**0.417**	0.261	**0.496**	0.322	0.383
	D4	0.397	0.038	0.141	**0.449**	0.529	0.5	0.514	**0.55**
	D5	0.373	**0.458**	0.254	0.305	0.472	0.472	0.264	**0.566**

TABLE V: Friedman Rank Sum Test on the Performance Measures of the Four Classification Methods, p is the p value and χ^2 is the chi square test statistic, $p < 0.05$ means there is a significant performance difference among the four methods.

	Γ_{ext}				Γ_{end}			
Measure	6 Hour		12 Hour		6 Hour		12 Hour	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p
F Score	4.92	0.177	10.92	**0.0121**	3.7826	0.286	7.8	0.050
Precision	1.08	0.781	12.12	**0.0069**	1.1739	0.759	7.8	0.050
Recall	8.1875	**0.0423**	9.72	**0.021**	7.6957	0.053	4.4667	0.215

TABLE VI: Nemenyi Post-hoc Test on the Performance Measures of the Four Classification Methods with 12 Hour Window Forecasting, the p value is recorded, $p < 0.05$ means the RVFL network significantly outperforms the benchmark methods(ANN, RF or SVM).

		Γ_{ext}			Γ_{end}		
		ANN	RF	SVM	ANN	RF	SVM
RVFL	F Score	0.203	0.961	**0.017**	0.20	0.99	0.12
	Precision	0.068	0.995	0.122	0.32	0.99	0.20
	Recall	0.122	0.611	**0.017**	0.53	0.69	0.20

C. Computation Time

The average and standard deviation of the computation time of training and testing over 5 datasets are shown in Table VII.

ANN has the longest training time among them all due to the iteration of BP. Because RF is regression tree based, the training and testing are fast. SVM requires grid search during training thus it has slow training speed but fast testing speed. RVFL has the advantage of fast training because it utilizes least square method to calculate the output layer weights as well as the fast testing because of the matrix calculation. From the table, we can see that in general RVFL has the best testing time yet comparable training time compared with RF.

TABLE VII: Computation Time (sec) of the Classification Methods over 5 Datasets, 'Ave': average/mean, and 'SD': standard deviation

Dataset	RF		SVM		ANN		RVFL	
	Train	Test	Train	Test	Train	Test	Train	Test
Γ_{ext}, Horizon=6 h								
D1	1.2	1.13	85.85	0.24	205.42	7.55	1.72	0.09
D2	0.88	0.9	75.56	0.2	176.92	4.52	1.7	0.09
D3	1.17	1.22	88.48	0.22	181.02	6.27	1.67	0.88
D4	0.78	0.81	72.88	0.21	157.63	4.48	1.89	0.08
D5	0.98	1.01	92.18	0.33	187.18	5.04	1.73	0.1
Ave	1	1	82.99	0.24	181.63	5.57	1.74	0.25
SD	0.182	0.166	8.37	0.05	17.29	1.32	0.09	0.35
Γ_{end}, Horizon=6 h								
D1	1.06	1.07	86.03	0.25	214.61	5.55	1.72	0.09
D2	0.92	1.12	78.76	0.12	173.92	5.48	1.75	0.08
D3	1.14	1.17	89.94	0.25	183.11	5.3	1.76	0.09
D4	1	0.78	71.74	0.21	157.38	4.07	1.81	0.08
D5	1	1.14	93.96	0.3	176.78	4.92	1.73	0.09
Ave	1.03	1.06	84.09	0.23	181.16	5.06	1.75	0.09
SD	0.08	0.16	8.89	0.07	20.97	0.61	0.04	0.01
Γ_{ext}, Horizon=12 h								
D1	0.81	0.78	64.36	0.22	151.3	3.95	1.37	0.08
D2	0.7	0.67	50.3	0.17	118.95	3.62	1.27	0.06
D3	0.67	0.7	57.47	0.16	113.88	4.57	1.22	0.06
D4	0.62	0.63	51.01	0.15	122.21	3.71	1.26	0.06
D5	0.89	0.85	64.46	0.2	141.37	3.32	1.39	0.08
Ave	0.74	0.73	57.52	0.18	129.54	3.83	1.30	0.07
SD	0.11	0.09	6.88	0.03	16.00	0.47	0.07	0.01
Γ_{end}, Horizon=12 h								
D1	0.98	0.8	70.18	0.21	154.19	4.42	1.55	0.08
D2	0.66	0.63	46.34	0.13	115.34	4.64	1.22	0.09
D3	0.78	0.98	64.57	0.2	139.76	4.15	1.61	0.05
D4	0.64	0.61	51.28	0.16	124.77	3.2	1.26	0.05
D5	0.85	1.09	71	0.21	146.03	3.52	1.48	0.08
Ave	0.78	0.82	60.67	0.18	136.02	3.99	1.42	0.07
SD	0.14	0.31	11.25	0.04	15.80	0.61	0.17	0.02

VI. Conclusion

This paper has covered an important topic on wind power utilization: wind power ramp forecasting. The importance and challenges of accurate wind power ramp forecasting have been addressed. Various wind power ramps have been defined in this paper. A random vector functional link (RVFL) neural network has been employed to forecast the wind power ramp and the ramp rate. The RVFL network has comparable performance as the benchmark methods: artificial neural network (ANN), random forests (RF) and support vector machine (SVM) for 6 hour window forecasting but RVFL network has better performance than the SVM for 12 hour window forecasting. The computation time of RVFL network has significant advantage over SVM and ANN and is comparable as RF.

Future work may includes to extend the power ramp classification from binary class to multiple class. The additional classes may refer to up-ramp and down-ramp events, strong ramp and weak ramp events, and fluctuation ramp events (ramp with more than one local maxima and minima).

Another possible future work is to switch from point forecasting to probabilistic forecast i.e. each forecast value is a set of probabilities and a corresponding interval instead of a single value.

Acknowledgement

This work was supported by the Singapore National Research Foundation (NRF) under its Campus for Research Excellence And Technological Enterprise (CREATE) programme, and Cambridge Advanced Research Centre in Energy Efficiency in Singapore (CARES), C4T project.

The author Ye Ren would like to thank the Clean Energy Program Office (CEPO) for scholarship.

References

[1] T. Ouyang, X. Zha, and L. Qin, "A survey of wind power ramp forecasting," *Energy and Power Engineering*, vol. 5, pp. 368–372, Jan. 2013.

[2] H. Zareipour, D. Huang, and W. Rosehart, "Wind power ramp events classification and forecasting: A data mining approach," in *Proc. IEEE Power and Energy Society General Meeting*. IEEE, 2011, pp. 1–3.

[3] M. Cui, D. Ke, Y. Sun, D. Gan, J. Zhang, and B.-M. Hodge, "Wind power ramp event forecasting using a stochastic scenario generation method," *IEEE Trans.Sustainable Energy*, vol. 6, no. 2, pp. 422–433, 2015.

[4] N. Francis, "Predicting sudden changes in wind power generation," *North American Windpower*, vol. 5, pp. 58–60, 2008.

[5] R. Girard, A. Bossavy, and G. Kariniotakis, "Forecasting ramps of wind power production at different time scales," in *European Wind Energy Conference*, 2011.

[6] J. Zhang, A. Florita, B. M. Hodge, and J. Freedman, "Ramp forecasting performance from improved short-term wind power forecasting," National Renewable Energy Laboratory (NREL), Tech. Rep., May 2014, nREL/CP-5D00-61730.

[7] H. Zheng and A. Kusiak, "Prediction of wind farm power ramp rates: A data-mining approach," *Journal of solar energy engineering*, vol. 131, no. 3, pp. 031 011–1–031 011–8, 2009.

[8] A. Bossavy, R. Girard, and G. Kariniotakis, "Forecasting uncertainty related to ramps of wind power production," in *European Wind Energy Conference and Exhibition 2010, EWEC 2010*, vol. 2. European Wind Energy Association, 2010, pp. 1–6, hal-00812403.

[9] ——, "A novel methodology for comparison of different wind power ramp characterization approaches," in *EWEA 2013-European Wind Energy Association annual event*, 2013, pp. 1–6.

[10] C. Ferreira, J. Gama, L. Matias, A. Botterud, and J. Wang, "A survey on wind power ramp forecasting," Argonne National Laboratory (ANL), Tech. Rep., 2010, aNL/DIS-10-13.

[11] G. Venayagamoorthy, K. Rohrig, and I. Erlich, "One step ahead: Short-term wind power forecasting and intelligent predictive control based on data analytics," *IEEE Power and Energy Magazine*, vol. 10, no. 5, pp. 70–78, Sep. 2012.

[12] R. L. Welch, S. M. Ruffing, and G. K. Venayagamoorthy, "Comparison of feedforward and feedback neural network architectures for short term wind speed prediction," in *Proc. International Joint Conference on Neural Networks (IJCNN2009)*, Jun. 2009, pp. 3335–3340.

[13] Y.-H. Pao, G.-H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," *Neurocomputing*, vol. 6, no. 2, pp. 163–180, 1994.

[14] C. L. P. Chen, "A rapid supervised learning neural network for function interpolation and approximation," *IEEE Transactions on Neural Networks*, vol. 7, no. 5, pp. 1220–1230, Sep. 1996.

[15] C. P. Chen and J. Z. Wan, "A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 1, pp. 62–72, 1999.
[16] (2015, Apr.) NREL west wind. NREL. [Online]. Available: wind.nrel.gov/Web_nrel/
[17] S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium (NAPS2010)*, Arlington, TX, 26-28 Sep 2010, pp. 1–8.
[18] P. Rosenmai. (2013, Nov.) Using the median absolute deviation to find outliers. Eurekastatistics. [Online]. Available: eurekastatistics.com/using-the-median-absolute-deviation-to-find-outliers
[19] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," in *Proc.Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2046. The Royal Society, 2004, pp. 1597–1611.
[20] Z. Wu and N. E. . Huang, *Hilbert-Huang transform and its applications*. World Scientific, 2005, vol. 5, ch. Statistical Significance Test of Intrinsic Mode Functions, pp. 107–125.
[21] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis," *Proc. Royal Society London A*, vol. 454, pp. 903–995, 1998.
[22] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sept 2009.
[23] C.-N. Ko and C.-M. Lee, "Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended kalman filter," *Energy*, vol. 49, pp. 413–422, 2013.
[24] W. F. Schmidt, M. A. Kraaijveld, and R. P. Duin, "Feedforward neural networks with random weights," in *Proc. IAPR International Conference on Pattern Recognition Conference B: Pattern Recognition Methodology and Systems*. IEEE, 1992, pp. 1–4.
[25] (2015, Apr.) Wind power generation data. ELIA. [Online]. Available: www.elia.be/en/grid-data/power-generation/wind-power