

## Research on Short-Term Wind Power Forecasting Based on RMADE-SEN-IRFR

Baoyi Wang<sup>a</sup>, Hui Wang<sup>b</sup>, Shaomin Zhang<sup>c</sup>

School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, China

<sup>a</sup>email:wangbaoyiqj@126.com, <sup>b</sup>email:ncepuwanghui@foxmail.com,

<sup>c</sup>email:zhangshaomin@126.com

**Keywords:** Wind Power Forecasting; Random Forest Regression; M5P Model Tree; Differential Evolution; Selective Ensemble

**Abstract.** To solve the problem of slowly convergence rate, weakly generalization ability, complicated model structure and parameter determination method of traditional machine learning model, the RFR model is introduced to forecast the short-term wind power. The wind power prediction model of RFR has the advantages of generalized error control, fast convergence rate, few adjustment parameters, strong interpretability, but its prediction accuracy is low and the computational complexity is high. So this paper constructs an RFR model using the M5P model tree and improves the accuracy of wind power prediction and reduces the computational complexity by combining the selective ensemble method (SEN) based on improved differential evolution algorithm (RMADE). Experimental results show that the proposed method has higher prediction accuracy and better generalization performance than SVM improved by GA (GA-SVM), OS-ELM and traditional RFR.

### Introduction

With the rapid growth of installed capacity of Wind Power (WP), the intermittence, randomness and volatility of wind power become more and more important for the stability and economy of power system [1]. Timely and accurate prediction of wind power can provide an important basis for rational management of the power grid, and improve the utilization of wind power effectively. In this paper, the active power prediction of wind generator in the future 24h is studied, and the time resolution is 10min.

### Related Works

In recent years, statistical learning method based on artificial intelligence has become a research hotspot of short-term wind power forecasting. In the paper [3, 4], SVM model is used to predict the short-term wind power and the niche genetic algorithm and the improved artificial fish swarm algorithm are used to optimize the parameters. However, the accuracy of the results depends on the accuracy and completeness of the selected kernel function and related parameters. In the paper [5,6], the radial basis function (RBF) and the online sequence limit learning machine (OS-ELM) model is used to predict the short-term wind power. But the RBF model is slow in training and easy to overfit, and the OS-ELM model has poor interpretability.

Random Forest Regression (RFR) is an integrated learning method with the advantages of high prediction precision, controllable generalization error, fast convergence rate, less regulation parameter and good interpretability [7]. Firstly, traditional RFR algorithm constructed by CART which calculate the single tree prediction value by the simple average method and its prediction precision of the meta-decision tree is low. In this paper, the M5P model tree is used to improve the RFR model. Secondly, a selective ensemble method based on random mutation adaptive evolutionary algorithm (RMADE) is adopted to delete some meta-decision trees with poor prediction performance and low diversity in RFR model.

### Analysis of Influencing Factors of Short-Term Wind Power

The environmental factors that affect the power of the wind generator are mainly wind speed ( $WS$ ), wind direction ( $WD$ ), air temperature ( $T$ ), air pressure ( $AP$ ), relative humidity ( $RH$ ). In this paper, we use the historical data collected from SCADA system of a wind farm in central Inner Mongolia to compute the Person correlation coefficient and Spearman Rank correlation coefficient of environmental factors and actual wind power. The results are shown in Table 1.

Table 1 The correlation coefficient between environmental factors and actual wind power

	$WS$	$WD$	$T$	$AP$	$RH$
Pearson	0.6846	0.0454	0.1968	-0.0843	-0.3057
Spearman Rank	0.8324	-0.0080	0.1637	-0.0734	-0.3490

According to Table 1,  $WS$ ,  $T$  and  $RH$  were selected when constructing the required sample attributes for the prediction model.

### An Improved RFR Model Based on M5P Model Tree

Random Forest Regression (RFR) is a random forest algorithm applied to regression problems. The main methods to improve the prediction accuracy of the RFR model include reducing the relevance of the meta-decision trees and improving the predictive performance of the meta-decision trees [7]. The traditional RFR algorithm uses CART as the meta-decision tree, and its leaf node sample prediction value is calculated by averaging method to calculate the final prediction result [7,8]. It can not give a continuous output, and the local leaf nodes are prone to be affected by the specific noise data. And there is a problem that the prediction error is large and the prediction result can not go beyond the historical prediction range. In this paper, we use the M5P model tree as the meta-decision tree to optimize the RFR algorithm.

M5P model tree is a segmented multiple linear regression tree proposed by Quinlan in 1992, where the leaf node is a multiple linear regression model [9]. The basic structure of M5P model tree is shown in Fig. 1. In the figure,  $LM_i, i = 1, 2, 3, 4$  is a multivariate linear regression model based on leaf node samples.

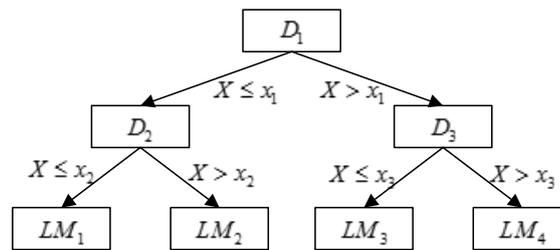


Fig. 1 The basic structure of M5P model tree

When the M5P model tree is generated, the criterion of dividing the attribute values of the nodes usually adopts standard deviation (SD) reduction maximization criterion [7]. The stopping criteria for node partitioning is the SD reduction percentage is less than the specified threshold ( $\gamma$ ) or the number of nodes after partitioning is less than the specified threshold ( $\sigma$ ).

$$SDR = SD_{D_k} - \frac{|D_{k_1}|}{|D_k|} SD_{D_{k_1}} - \frac{|D_{k_2}|}{|D_k|} SD_{D_{k_2}} \tag{1}$$

Wherein,  $SDR$  is the SD reduction value after dividing the node  $N_k$  according to the attribute value  $x_i$ ,  $SD_{D_k}$ ,  $SD_{D_{k_1}}$ ,  $SD_{D_{k_2}}$  are the SD of sample space  $D_k$  and subspace  $D_{k_1}$ ,  $D_{k_2}$ , respectively.

### Improved RFR Prediction Model Based on RMADE-SEN

The selective ensemble method [11] is proved that it can be constructed better than the use of all the basic learners by choosing partial basis learners [11]. In this paper, Differential evolution algorithm

(DE) is introduced into selective ensemble of M5P-RFR model to forecast the short-term wind power. DE is a kind of random search algorithm based on swarm intelligence, which can memorize the individual vector optimal solution and share the information within the population [11]. Standard DE algorithm parameters of the scaling factor  $F$ , crossover probability  $CR$  is fixed, it is difficult to balance the global search and local search. In paper [13, 14], a parameter adaptive distributed DE algorithm (APDDE) and a random mutation DE algorithm (RMDE) are proposed. A DE algorithm with bidirectional random multi-strategy mutation and parameter adaptation is proposed in paper [15]. In this paper, a multi-strategy random mutation and parameter adaptive differential evolution algorithm is proposed, which is based on evolution generation to adjust global optimization and local search ability automatically.

**Random Mutation Operation Based on Multi-Strategy.** By analyzing the characteristics of four commonly used differential mutation operations [15], a random mutation operation based on multi-strategy is proposed in this paper, which can adaptively adjust the global optimization and the local search ability according to the evolution generation. The expression is as follows:

$$v_{i,g} = \begin{cases} x_{r_0,g} + F_1 \cdot (x_{r_1,g} - x_{r_2,g}), & \text{if } rand_1 \leq 0.5 \text{ and } rand_2 \leq t \\ x_{best,g} + F_2 \cdot (x_{r_1,g} - x_{r_2,g}), & \text{if } rand_1 \leq 0.5 \text{ and } rand_2 > t \\ x_{i,g} + F_{31} \cdot (x_{r_0,g} - x_{i,g}) + F_{32} \cdot (x_{r_1,g} - x_{r_2,g}), & \text{if } rand_1 > 0.5 \text{ and } rand_2 \leq t \\ x_{i,g} + F_{41} \cdot (x_{best,g} - x_{i,g}) + F_{42} \cdot (x_{r_1,g} - x_{r_2,g}), & \text{if } rand_1 > 0.5 \text{ and } rand_2 > t \end{cases} \quad (2)$$

Where,  $rand_1$  and  $rand_2$  is a random number of uniform distribution in interval  $[0,1]$  and  $[1,T]$ , respectively.  $t$  is the current evolutionary generation of the population,  $T$  is the largest evolutionary generation of the population,  $F$  is the scale factor.

**Adaptive Selection of Scale Factor.** Different variation operations have different requirements for the change of the scaling factor. In this paper, different scaling factors are used for different mutation operations. The method is as follows:

$$\begin{cases} F_1 = F_{31} = F_{32} = F_{42} = \frac{t-1}{T-1} \cdot (F_{max} - F_{min}) + F_{min} \\ F_2 = F_{41} = \frac{t-1}{T-1} \cdot (F_{min} - F_{max}) + F_{max} \end{cases} \quad (3)$$

Wherein,  $F_{max}$  and  $F_{min}$  are the upper and lower limits of the scaling factor, respectively.

**Evaluation of RFR Model.** This paper presents a comprehensive evaluation method which based on the comprehensive measure of RFR algorithm's generalization error, prediction error and meta-decision tree error distribution. The formula for evaluating the fitness of the IRFR model  $f(w)$  is as follows:

$$f(w) = OOB(w) + RMSE(w) - RMSE_{rmse}(w) \quad (4)$$

$$OOB(w) = \frac{\sum_{i=1}^K w_i OOB_i}{\sum_{j=1}^K w_j} \quad (5)$$

$$RMSE(w) = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (P_{Mi} - P_i(w))^2} \quad (6)$$

$$RMSE_{rmse}(w) = \sqrt{\frac{1}{K} \sum_{j=1}^K (RMSE_j - RMSE(w))^2} \quad (7)$$

Where,  $OOB_i$  is generalization error of the  $i$  th meta-decision tree OOB samples,  $w$  is the weight vector of the meta-decision tree;  $OOB(w)$ ,  $P_i(w)$  and  $RMSE(w)$  is the generalization error, the final prediction and the prediction error of the model, respectively;  $RMSE_j$  is the error prediction of a single meta-decision tree,  $RMSE_{rmse}(w)$  is the discrete degree of the prediction error of the meta-decision tree in the IRFR model.

When  $OOB(w)$  is smaller,  $RMSE(w)$  is smaller and  $RMSE_{mse}(w)$  is bigger, the model's fitness value  $f(w)$  is smaller, and the model forecast performance is better.

**Strategy of Selective Ensemble.** Assume that the optimal weight vector of the RFR model is  $w^* = \{w_1^*, w_2^*, \dots, w_K^*\}$ . Let  $\lambda$  be the weight threshold of selective ensemble. If  $w_i^* / \bar{w}^* \geq \lambda, i = 1, 2, \dots, K$  is satisfied, where  $\bar{w}^*$  is the average of weight, the corresponding meta-decision tree of  $w_i^*$  is retained, otherwise it is deleted from the RFR model; Finally, we obtain  $S$  decision tree  $\{T_1, T_2, \dots, T_S\}$  with the corresponding weight vector  $\{w_1^*, w_2^*, \dots, w_S^*\}$ , which constitutes the optimized RFR model.

### Forecasting Steps of Wind Power Based on RMADE-SEN-IRFR

Let  $D$  be the training set of historical wind power monitoring data,  $N$  be the number of training set samples,  $M$  be the training set sample attribute number,  $K$  be the number of M5P meta-decision trees.  $\lambda$  be the selective ensemble weight threshold,  $NP$  be the population size of the RMADE,  $T$  be the largest evolutionary generation of the population,  $[F_{min}, F_{max}]$  be the scale factor value range,  $[CR_{min}, CR_{max}]$  be the crossover probability interval. The basic steps of short-term wind power forecasting based on RMADE-SEN-IRFR is as follows:

(1) Resampling from training set  $D$  by Bootstrap to obtain a new training set  $D_i$  which is equal to the number of the original training set as the current meta-decision tree root node sample space.

(2) If the number of current node samples is less than  $\sigma$ , stop the partition, mark the current node as the termination node, go to Step (4). Otherwise,  $m(m \ll M)$  attributes are randomly selected from  $M$  attributes, the sample space is partitioned into two subspaces according to the possible attribute values, and  $SDR$  is calculated according to formula (1) to get the optimal partition attribute value.

(3) If the  $SD$  reduction percentage of the optimal partition is less than the threshold  $\gamma$ , stop the partition, mark the current node as the termination node, and go to Step (4). Otherwise, divide the current node into two sub-nodes according to the optimal partition.

(4) If all the leaf nodes in the current meta-decision tree are termination nodes, go to Step (5). Otherwise, take any leaf node of non-terminating node, go to Step (2) and continue the division.

(5) Traversing all leaf nodes of the current meta-decision tree and constructing a multiple linear regression model at each leaf node.

(6) Repeating Step (1) ~ Step (5), get  $K$  M5P meta-decision trees, constitute the IRFR model based on M5P.

(7) Generating the initial population  $W = \{w^0, w^1, w^2, \dots, w^{NP}\}$  is randomly by the real value coding method. Wherein,  $w^0$  is the reference weight vector from the meta-decision tree's OOB error,  $w^i$  is a weight vector generated randomly. Let the current evolutionary generation be  $t = 1$ .

(8) Calculating the individual fitness value, complete the differential mutation, crossover, select the operation, and then generating the next generation of individual population, setting  $t = t + 1$ . If  $t > T$ , then stop evolution, otherwise repeat Step (8).

(9) Let the vector of the optimal individual weight be  $w^* = \{w_1^*, w_2^*, \dots, w_K^*\}$ . According to the selective integration strategy,  $S$  meta-decision trees  $\{T_1, T_2, \dots, T_S\}$  are preserved, and the corresponding weight vector is  $\{w_1^*, w_2^*, \dots, w_S^*\}$ .

(10) Constructing a sample of test set and calculating the weighted average of  $S$  meta-decision trees as the final prediction result. Adopting the multi-step forecasting method to predict the wind power value in the future 24h.

### Experiment and Result Analysis

The experimental data are from SCADA historical monitoring data and NWP data of a wind farm in Inner Mongolia in February 2016. The data from February 8 to February 24 are used as the training

set and the data from February 25 to February 27 as the test set. The multi-step prediction method is used to predict the wind power of 144 points in the future. The root mean square error (RMSE) and the average absolute error (MAE) are selected as the evaluation criteria of the RMADE-SEN-IRFR model forecasting effect. The formula is as follows:

$$RMSE = \frac{\sqrt{\sum_{i=1}^N (P_{Mi} - P_i)^2}}{P_{cap} \cdot \sqrt{N}} \times 100\% \quad (10)$$

$$MAE = \frac{\sum_{i=1}^N |P_{Mi} - P_i|}{P_{cap} \cdot N} \times 100\% \quad (11)$$

Wherein,  $P_{Mi}$  and  $P_i$  are the actual power and the predicted power at time  $i$ ,  $P_{cap}$  is the average boot capacity,  $N$  is the number of samples in the test sample set.

After the experiment debugging, set the parameter  $K = 200$ ,  $m = \lfloor M / 3 \rfloor = 94$ ,  $\sigma = 10$ ,  $\gamma = 5\%$ ,  $NP = 100$ ,  $T = 50$ ,  $[F_{min}, F_{max}] = [0.3, 0.95]$ ,  $[CR_{min}, CR_{max}] = [0.1, 0.9]$  and  $\lambda = 0.8$ .

Forecast the short-term wind power of the next 24 hours from February 25 to February 27, by using the proposed RMADE-SEN-IRFR model, GA-SVM [2], OS-ELM [4] and the traditional RFR algorithm. The forecasted error is averaged 30 times. According to Table 2, Fig. 1 and Fig. 2, we can see that the algorithm RMADE-SEN-IRFR adopted in this paper is superior to GA-SVM algorithm, OS-ELM algorithm and traditional RFR algorithm. The traditional RFR algorithm has larger prediction error, OS-ELM algorithm has smaller prediction error but the forecast value fluctuates greatly.

Table 2 Wind power prediction error of four algorithms

Algorithm	RMSE(%)	MAE(%)
RFR	15.67	14.63
RMADE-SEN-IRFR	8.94	6.91
GA-SVM [2]	9.37	7.89
OS-ELM [4]	11.87	9.40

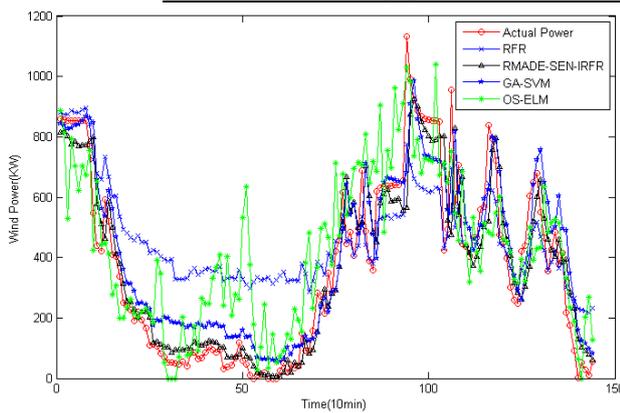


Fig. 2 Comparison of actual wind and the predicted results

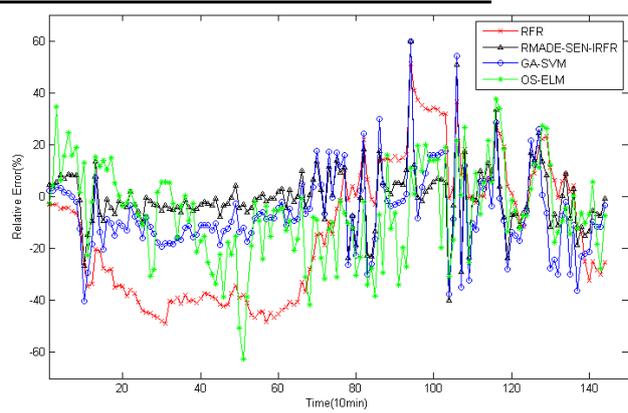


Fig. 3 The relative error of the four algorithms

## Conclusion

In this paper, Improved RFR model based on M5P model tree is applied to short-term wind power forecasting. The redundant meta-decision tree in the RFR model is removed by the RMADE-SEN method, which improved by combining multi-strategy random mutation and parameter adaptive method and comprehensive evaluation method of model. The experimental results show that the short-term wind power prediction model RMADE-SEN-IRFR proposed in this paper has better performance.

## **Acknowledgement**

In this paper, the research was sponsored by the National Natural Science Foundation of China (Project No. 61300040) and Scientific Research Project of Hebei Province (Project No. Z2012077 ).

## **Literature References**

- [1] QIAN Zheng, PEI Yan, CAO Lixiao, et al, Overview of Wind Power Prediction Methods, *J. High Voltage Technology*, 2016, (04): 1047-1060.
- [2] Xue Yusheng, Xu Chen, et al, Comments on short - term and ultra - short term wind power prediction, *J. Automation of Electric Power Systems*, 2015, 06: 141-151.
- [3] Yan Xiaojuan, Gong Renxi, Short-term wind power forecasting based on SVM optimized by Genetic Algorithm, *J. Electric Measurement and Instrumentation*, 2014,08: 38-41 + 46.
- [4] Wang Yongxiang, Chen Guochu, Short-term wind power prediction based on SVM optimized by fish swarm algorithm, *J. Electric Measurement and Instrumentation*, ,2016,03:80-84.
- [5] Ma Bin, Zhang Liyan, A method of short-term wind power direct prediction based on radial basis function neural network, *J. Power System Protection and Control*, 2015, (19): 78-82.
- [6] WANG Yan, WANG Zhen, et al, Super short-term wind power prediction based on OS-ELM and Bootstrap method, *J. Automation of Electric Power Systems*, 2014,(06):14-19+122.
- [7] Wu Xiaoyu, He Jinghan, et al, Short-term load forecasting of power system based on random forest algorithm improved by gray projection, *J. Automation of Electric Power Systems*, 2015,12: 50-55.
- [8] WANG Dwen, SUN Zhiwei, General data analysis and parallel load forecasting of power users, *J. Transactions of China Electrotechnical Society*, 2015, (03): 527-537.
- [9] Zhang Jianmin, Liu Dengtao, et al, Using M5P model tree and measurement data to identify steam consumption characteristics of extraction steam turbine, *J. Transactions of China Electrotechnical Society*, 2011,(23):21-26.
- [10] GUO Qi, CHEN Tianshi, CHEN Yunji, Multi-core design space exploration technology based on model tree, *J. Journal of Computer-Aided Design and Computer Graphics*, 2012, (06): 710-720.
- [11] ZHANG Chunxia, ZHANG Shishe, A Survey of Selective Integrated Learning Algorithms, *J. Chinese Journal of Computers*, 2011, (08): 1399-1410.
- [12] ZHANG Mingqing, DI Fuqiang, LIU Jia, General steganalysis based on selective ensemble classifier, *J. Journal of Sichuan University (Engineering Science Edition)*, 2015, (01): 36-41.
- [13] ZHANG Chunmei, CHEN Jie, XIN Bin, Parameter adaptive distributed differential evolution algorithm, *J. Control and Decision*, 2014, (04): 701-706.
- [14] OUYANG Haibin, GAO Liqun, et al, Differential evolutionary algorithm of random mutation, *J. Journal of Northeastern University (Natural Science Edition)*, 2013, (03): 330-334.
- [15] KONG Xiangyong, GAO Liqun, et al, Adaptive differential evolution algorithm for random multi-strategy mutation, *J. Computer Integrated Manufacturing Systems*, 2014, (08): 1948-1958.