Research on Reduced Scene Sets Based on ARMA Model of Wind Farms Day-ahead Total Output Forecasting

Hui chao WANG¹, Lei LIU², Jin hui MA³, Jie DING¹ and Chang ZHOU¹

¹Renewable Energy Research Center, China Electric Power Research Institute, Nanjing, 210003, China
²Power Dispatch Department, State Grid Jiangxi Electric Power Company, Jiangxi, 330077, China
³Power Dispatch Department, State Grid Anhui Electric Power Company, Anhui, 230061, China
E-mail: 17600271330@163.com

Abstract: In this paper, the ARMA model is used to linearly fit the time series data of wind farms’ prediction error with the software E-views. And then the error linear curve is sampled by four sampling methods, including random sampling, important sampling method, Latin hypercube sampling method and quasi-Monte Carlo method, to obtain some Ascending and disorderly samples respectively. Finally, the reduced scene sets are obtained by substituting the samples into the scene reduction model. Through analysing the reduced scenes output curve with the evaluation indicators of wind farms’ forecast output curve, we find that the reduced scenarios are closer to the actual output curve than the traditional predicted. It also can be concluded that they have great effect on prediction correction and sampling methods have little effect on the output trend of reduced scenarios. Whereas, comparing the reduced scenes’ output curve before and after sorting the sample data, the disorder and randomness of the sample data will lead to great volatility in the reduced scenes.

1. Introduction

With the continuous increase of installed capacity and output contribution of grid-connected wind farms, the uncertainty of wind farms’ output prediction will pose a significant threat to safe and stable operation of power system. To deal with this problem, we often use the scene analysis technique that converts the uncertainty to some deterministic scenarios with unequal probability.

At present, the ARMA model is commonly used to analyze and model the forecasting data and combined with the sampling methods and the current prediction output to generate the basic scene sets. In order to reduce the calculation amount and time when optimizing the generation plan, we often use the scene reduction technology to reduce the number of basic scenes and get the reduced scene sets [1]. In this process, the commonly used sampling methods are the random sampling method [2]-[3], the important sampling method, the Latin hypercube sampling method, and the quasi-Monte Carlo method [4]-[6].

2. ARMA model

The ARMA model, also known as the Auto-Regressor and Moving Average model, is the most extensive time series model. It not only reveals the structure and regularity of dynamic data,
quantitatively understands the linear correlation of observed data and predicts its future value, but also can study the relevant characteristics of the system from various aspects. In this paper, the best ARMA model will be used to linearly fit the historical forecasting error data of wind farms with the E-views software. The model expression is:

$$\Delta p^i_g(t) = \sum_{i=0}^{p} \alpha_i \Delta p^i_g(j) - \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t \quad (\Delta p^i_g = 0, \varepsilon_t = 0, t <= 0)$$

(1)

Where, $p, q$ are the order of ARMA model, $\alpha_i, \beta_j$ are the model parameters, $\varepsilon$ is a Gaussian distribution with mean 0 and variance $\sigma^2$, $\Delta p^i_g$ is the wind farm forecast error rate at time $t$.

3. Reduced scene set

3.1. Basic scene set

The reduced scene sets are derived from the base scene sets with the scene reduction technology. And the basic scenes are generated by ARMA model with various sampling methods. Sampling methods to sample $\varepsilon$, the sample is $T$ data, namely $\varepsilon_1(s), \ldots \varepsilon_T(s)$, and then substituted into equation (1) to obtain the prediction error rate $\Delta p^i_g(s)$ at each time point. Substituting them into equation (2) to get the output forecast scene of wind farm.

$$p^i_g(s) = p^i_g - C \times \Delta p^i_g$$

(2)

Where, $p^i_g$ is forecasting power of wind farm at time $t$, $C$ is the installed capacity of the wind farm.

The above steps are repeated to generate $N$ forecasting output scenes with probability of $1/N$ to form a basic scene set.

3.2. Reduce scenes

In order to reduce the amount and time of calculation when using the scene analysis technique to optimize the generation schedule, the scene reduction technique will be used to reduce the number of basic scenes. The basic idea of the scene reduction method is to minimize the possible distance between the basic scene set and the subset of scenes that are ultimately reserved. That is, under the condition given by the number of deleted scenes $J$, the pursuit of formula (3) is the smallest.

$$\min_{J \subseteq \mathcal{S}} \left\| p^i_g - p^i_g \right\|_2$$

(3)

Where, $\mathcal{J}$ is the set of elimination scenes in the scene reduction process. This paper uses the backward reduction method to reduce the basic scene set.

3.3. Sampling method

At present, the classical sampling methods are random sampling, important sampling, Latin hypercube sampling and quasi-Monte Carlo [7]. Random sampling aims to ensure that each value of the variable may have a known, non-zero probability to be sampled to ensure sample representativeness. Whereas, the important sampling method will modify the given probability distribution function before sampling to make the sample more subjective. The paper uses the kernel function method to modify the original probability distribution function.

The Latin hypercube method and the quasi-Monte Carlo method both improve the sample generation process and effectively reflect the overall distribution of variables, ensuring that all sample areas can be covered by sampling points. Differently, the Latin hypercube sampling method is an equal probability interval sampling while the quasi-Monte Carlo method is a low differential sequence sampling.

4. Wind farm forecast evaluation indicators

Wind farm forecasting and evaluation indicators can accurately measure wind farm forecasting errors, which mainly including maximum error rate, accuracy rate, qualification rate, correlation coefficient and comprehensive evaluation index [8].
The maximum error rate is used to describe the maximum error data in all time points. The expression is:

$$EV = \max \left( \frac{|P_f - P_m|}{C_i} \right) \times 100\% \quad (4)$$

Where, $P_f$ is the predicted power at time $i$, $P_m$ is the measured power at time $i$, and $C_i$ is the startup capacity of wind farm at time $i$.

The accuracy rate is the overall error range describing the predicted curve and the actual output curve. The expression is:

$$CR = \left(1 - \sqrt{\frac{1}{2} \sum_{i=1}^{n} \left( \frac{P_f - P_m}{C_i} \right)^2} \right) \times 100\% \quad (5)$$

Where, $n$ is the number of time points in the predicted and actual output curve.

The qualified rate is the ratio of the number of qualified points in the prediction curve. The expression is:

$$QR = \frac{1}{n} \sum_{i=1}^{n} B_i \times 100\% \quad (6)$$

Where, $B_i$ is 1 when the prediction error is qualified at time $i$, and 0 otherwise.

The correlation coefficient is the degree of correlation between the forecasting output curve and the actual output curve of the wind farm. The expression is:

$$R = \frac{\sum_{i=1}^{n} (P_f - \overline{P_f})(P_m - \overline{P_m})}{\sqrt{\sum_{i=1}^{n} (P_f - \overline{P_f})^2} \sqrt{\sum_{i=1}^{n} (P_m - \overline{P_m})^2}} \times 100\% \quad (7)$$

Where, $\overline{P_f}$ is the average of the predicted output curve, $\overline{P_m}$ is the average of the actual output curve.

The comprehensive evaluation index is used to evaluate the overall accuracy of the forecasting curve and the combination of the above four indicators. The expression is:

$$CEI = k_1 \times (1 - EV) + k_2 \times CR + k_3 \times QR + k_4 \times R \quad (8)$$

$$k_1 + k_2 + k_3 + k_4 = 1 \quad (9)$$

Where, $k_1, k_2, k_3, k_4$ are weight coefficients corresponding to different error indicators and 0.25 in the paper.

5. Analysis of examples

The data of calculation example includes the predicted total output curve and actual total output curve of some wind farms on a certain day in the province (the total installed capacity of some wind farms is 1917.2MW) and the prediction error data of the day. The specific data is shown in Figure 1, the blue curve is the prediction error curve, the black curve is the predicted total output curve and the red curve is the actual total output curve. At the same time, the evaluation index data of the prediction scene is displayed in Table 1.

![Figure 1. Forecast data of some wind farms in the province in one day](image)
Table 1. The evaluation data of wind farm’s day-ahead forecast output curve

| Evaluation index | Maximum error rate | accuracy rate | Qualified rate | Correlation coefficient | comprehensive evaluation index |
|------------------|--------------------|---------------|----------------|-------------------------|--------------------------------|
| Prediction scenario | 0.1136             | 0.9673        | 1.0000         | 0.4242                  | 0.819                          |

The software E-views is used to calculate the ARMA model. The model parameters are used to solve the residual series, and then the approximate maximum likelihood estimation method is used to solve the white noise variance. The ARMA model of the prediction error data is ARMA (2,6), the formula is:

$$
\Delta p_t = 1.6 \Delta p_{t-1} - 0.8 \Delta p_{t-2} + 0.4 \varepsilon_{t-1} - 0.03 \varepsilon_{t-2} - 0.2 \varepsilon_{t-3} - 0.3 \varepsilon_{t-4} - 0.2 \varepsilon_{t-5} - 0.2 \varepsilon_{t-6} + \varepsilon_t \quad (10)
$$

Combining with four sampling methods and backward reduction method, two reduced scene sets are generated to depict the uncertainty of the prediction. And one set is obtained before sorted sample data, the other is obtained after sorted sample data in ascending order.

5.1. Random sampling

(1) The sample data is not sorted

![Figure 2. Reduced scenes of wind farms with random sampling](image)

Table 2. Evaluation parameters of reduced scenes with random sampling

| Reduce scenes | scenario 1 | scenario 2 | scenario 3 | Expected value |
|---------------|------------|------------|------------|----------------|
| Probability   | 0.225      | 0.297      | 0.478      | 1              |
| Maximum error rate | 0.131      | 0.113      | 0.090      | 0.106          |
| accuracy rate | 0.963      | 0.894      | 0.898      | 0.911          |
| Qualified rate | 1.000      | 1.000      | 1.000      | 1.000          |
| Correlation coefficient | 0.436      | 0.552      | 0.553      | 0.526          |
| comprehensive evaluation index | 0.817      | 0.804      | 0.811      | 0.833          |

(2) Ascending sorting of sample data
Figure 3. Reduced scenes of wind farms with random sampling

Table 3. Evaluation parameters of reduced scenes with random sampling

| Reduce scenes | scenario 1 | scenario 2 | scenario 3 | Expected value |
|---------------|------------|------------|------------|----------------|
| Probability   | 0.650      | 0.008      | 0.342      | 1              |
| Maximum error rate | 0.089  | 0.090      | 0.066      | 0.081          |
| accuracy rate | 0.968      | 0.895      | 0.900      | 0.944          |
| Qualified rate | 1.000      | 1.000      | 1.000      | 1.000          |
| Correlation coefficient | 0.638  | 0.633      | 0.738      | 0.672          |
| comprehensive evaluation index | 0.879  | 0.861      | 0.868      | 0.884          |

Random sample data varies greatly. When the sample data is not sorted, the general trend of the forecasting reduced scenes is same with the actual output curve, but they all have strong volatility which make the overall evaluation slightly better than the predicted scene. After sorting the sample data in ascending order, the overall trend of the reduced scenarios is not changed, but the prediction curve is less volatile and the overall evaluation has been greatly improved.

5.2. Important sampling method

(1) The sample data is not sorted

Figure 4. Reduced scenes of wind farms with the important sampling method

Table 4. Evaluation parameters of reduced scenes with the important sampling method

| Reduce scenes | scenario 1 | scenario 2 | scenario 3 | Expected value |
|---------------|------------|------------|------------|----------------|
| Probability   | 0.122      | 0.273      | 0.605      | 1              |
The important sampling method modifies the white noise probability distribution function using the kernel function method with a width of 1, so that the samples have some similarity with each other. Therefore, when the sample data is not sorted, the reduced scenarios have more similar volatility. Comparing the predicted output curve, the overall trend of the reduced scenes is closer to the actual output curve, but the curve fluctuates strongly. After sorting the sample data in ascending order, the fluctuation of the reduced scenarios is reduced, and the overall forecasting evaluation is more better.

5.3. Latin hypercube sampling
(1) The sample data is not sorted

The important sampling method modifies the white noise probability distribution function using the kernel function method with a width of 1, so that the samples have some similarity with each other. Therefore, when the sample data is not sorted, the reduced scenarios have more similar volatility. Comparing the predicted output curve, the overall trend of the reduced scenes is closer to the actual output curve, but the curve fluctuates strongly. After sorting the sample data in ascending order, the fluctuation of the reduced scenarios is reduced, and the overall forecasting evaluation is more better.
Table 6. Evaluation parameters of reduced scenes with Latin hypercube sampling

| Reduce scenes | scenario 1 | scenario 2 | scenario 3 | Expected value |
|---------------|------------|------------|------------|----------------|
| Probability   | 0.387      | 0.377      | 0.236      | 1              |
| Maximum error rate | 0.086      | 0.081      | 0.108      | 0.090          |
| accuracy rate  | 0.970      | 0.895      | 0.899      | 0.925          |
| Qualified rate | 1.000      | 1.000      | 1.000      | 1.000          |
| Correlation coefficient | 0.590103   | 0.569      | 0.592      | 0.583          |
| comprehensive evaluation index | 0.868      | 0.851      | 0.845      | 0.854          |

Figure 7. Reduced scenes of wind farms with Latin hypercube sampling

Table 7. Evaluation parameters of reduced scenes with Latin hypercube sampling

| Reduce scenes | scenario 1 | scenario 2 | scenario 3 | Expected value |
|---------------|------------|------------|------------|----------------|
| Probability   | 0.007      | 0.991      | 0.002      | 1              |
| Maximum error rate | 0.084      | 0.0837     | 0.092      | 0.084          |
| accuracy rate  | 0.969      | 0.895      | 0.898      | 0.895          |
| Qualified rate | 1.000      | 1.000      | 1.000      | 1.000          |
| Correlation coefficient | 0.666      | 0.667      | 0.663      | 0.667          |
| comprehensive evaluation index | 0.888      | 0.869      | 0.868      | 0.870          |

Latin hypercube sampling divides the white noise probability distribution function into equal probability intervals, and then sampling in each probability interval randomly. Sample values can effectively reflect the overall distribution of variables. When the sample is not sorted, the overall forecasting evaluation of the reduced scenarios is better than the predicted output scenario, but they have very strong volatility. After the samples are sorted in ascending order, there is strong similarity between the samples for the sampling space is same and small at the same time point, and the fluctuation of reduced scenes is reduced.

5.4. Monte Carlo sampling method

(1) The sample data is not sorted

Quasi-Monte Carlo sampling is a low differential sequence sampling, and the sample can effectively reflect the overall distribution of variables. When the sample is not sorted, the overall trend of reduced scenario is closer to the actual output than the predicted scenario, but they also have strong volatility. After sorting the sample data in ascending order, the output fluctuation of reduced scenes is greatly reduced, and the overall evaluation is more better.
Figure 8. Reduced scenes of wind farms with Monte Carlo sampling method

Table 8. Evaluation parameters of reduced scenes with Monte Carlo sampling method

| Reduce scenes | scenario 1 | scenario 2 | scenario 3 | Expected value |
|---------------|------------|------------|------------|----------------|
| Probability   | 0.300      | 0.430      | 0.270      | 1              |
| Maximum error rate | 0.095      | 0.095      | 0.075      | 0.090          |
| accuracy rate | 0.970      | 0.895      | 0.898      | 0.918          |
| Qualified rate | 1.000      | 1.000      | 1.000      | 1.000          |
| Correlation coefficient | 0.606      | 0.600      | 0.578      | 0.596          |
| comprehensive evaluation index | 0.870      | 0.852      | 0.857      | 0.856          |

(2) Ascending sorting of sample data

Figure 9. Reduced scenes of wind farms with Monte Carlo sampling method

Table 9. Evaluation parameters of reduced scenes with Monte Carlo sampling method

| Reduce scenes | scenario 1 | scenario 2 | scenario 3 | Expected value |
|---------------|------------|------------|------------|----------------|
| Probability   | 0.497      | 0.060      | 0.443      | 1              |
| Maximum error rate | 0.097      | 0.091      | 0.092      | 0.095          |
| accuracy rate | 0.966      | 0.895      | 0.898      | 0.932          |
| Qualified rate | 1.000      | 1.000      | 1.000      | 1.000          |
| Correlation coefficient | 0.624      | 0.623      | 0.631      | 0.627          |
| comprehensive evaluation index | 0.873      | 0.857      | 0.858      | 0.866          |

6. Conclusion
Based on the ARMA model, this paper linearly fits the wind farms’ day-ahead prediction error data and generates the reduced scene sets with four sampling methods. The following conclusions can be drawn from analysis, evaluation and comparison of these reduced scenes.

1) The reduced scenes are closer than the predicted output curve. It can be seen that wind farms predictive reduction scenes do have a great prediction correction.
2) The four sampling methods have little effect on the trend of the reduced scenes output curve, but the disorder and randomness of sample data will make the reduced scenes more volatile.
3) In order to get better prediction scenarios, it is necessary to sort sample data reasonably to reduce the fluctuation of reduced scenes.

7. References
[1] LEI Yu, YANG Ming, HAN XUE Shan. A two-stage stochastic optimization of unit commitment considering wind power based on scenario analysis [J]. Power System Protection and Control, 2012, 40(23): 58-67.
[2] DENG Wei, LI Xin-ran, XU Zhen Hua, et al. Calculation and impact analysis of probability of wind speed correlation [J]. Power System Technology, 2012, 36(04): 45-50.
[3] DING Ming, WU Xing-long, LU Wei, et al. Three-phase random flow calculation of distribution network with multiple asymmetric photovoltaic grid-connected systems [J]. Automation of Electric Power Systems, 2012, 36(16):47-52.
[4] B. Zou and Q. Xiao, "Solving Probabilistic Optimal Power Flow Problem Using Quasi Monte Carlo Method and Ninth-Order Polynomial Normal Transformation," in IEEE Transactions on Power Systems, vol. 29, no. 1, pp. 300-306, Jan. 2014.
[5] HAIJAN M, ROSEHART W D, ZAREIPOUR H. Probabilistic power flow by Monte Carlo simulation with Latin super cube sampling [J]. IEEE Trans on Power Systems, 2013, 28(2):1550 - 59.
[6] HUANG H Z, CHUNG C Y, CHAN K W, et al. Quasi -Monte Carlo based probabilistic small signal stability analysis for power system with plug-in electric vehicle and wind power integration [J]. IEEE Trans on Power Systems, 2013, 28(3): 3335 – 43.
[7] LIU Yu, GAO Shan, YANG Sheng Chun, et al. Overview of probabilistic trend algorithms for power systems [J]. Automation of Electric Power Systems, 2014, 38(23):127-35.
[8] THE National Energy Administration. NB/T 31046-2013 Function specification of wind power forecast system [S]. Beijing: Electric Power Press of China, 2014