A Novel Continuous Permutation Method for Wind Power Correlation Analysis

Quan Li, Student Member, IEEE, and Nan Zhao, Member, IEEE

Abstract—Wind speeds in geographically adjacent areas are highly correlated, which correspondingly leads to wind power correlation. It is essential to consider the wind power correlation for steady-state related calculations in the modern power system with high wind power penetration. This short article presents a novel continuous permutation method for wind power correlation analysis. With the novel eigen-permutation, the proposed method is applicable to tackle the correlation analysis even the correlation matrix of random variables is not positive definite. A continuous permutation is then proposed to reduce the permutation deviation and, hence, to reduce the error of steady-state related calculations. Simulation results show the effectiveness of the proposed method in dealing with the nonpositive definite correlation matrix of the wind speeds. Also, compared to the traditional methods, the proposed method achieves more than an 80% reduction of the general error in the correlation analysis, which also contributes to higher accuracy in the power flow calculation.

Index Terms—Correlation problem, eigen-decomposition, renewable energy, sample permutation.

I. INTRODUCTION

The output of renewable energy generations, represented by wind farms (WFS), is commonly influenced by the variability of natural conditions. In the modern power system with high wind power penetration, some WFSs are installed in geographically adjacent areas where the wind speeds are correlated. To ensure the accuracy of steady-state related calculations such as probabilistic power flow (PPF) calculation and steady-state security analysis, the correlation between wind power needs to be fully considered.

At present, the copula method and permutation method are the most commonly used ways to analyze the wind power correlation. The copula method constructs the multidimensional distribution function (MDF) by using copula functions [1], thus, correlated samples can be obtained from the MDF. Different kinds of copula functions, such as Gaussian copula [2] and D-vine copula [3], are introduced to model the correlated outputs of WFSs. However, the copula method heavily relies on the time-consuming MDF sampling. Besides, combinations of copula functions are required for precisely describing the correlation of random variables, causing the increase of the computational burden. The permutation method generates the sample matrix, which contains the correlation information by permuting the samples of each variable. Since the permutation method does not rely on the MDF, it has broader applications such as combining with Monte Carlo method [4]–[6] and point estimation method [7], which are widely used to analyze uncertainties in power systems. Also, there is no time-consuming MDF sampling in permutation methods. The Cholesky decomposition method, the most commonly used permutation method, is applied to permute the samples and generate the sample matrix with correlation information [4]–[9]. However, the Cholesky decomposition method is not applicable to the case that the correlation matrix of variables is positive definite. In addition, all aforementioned studies did not consider the deviation of the correlation matrix caused by the permutation, which causes errors in the correlation analysis and steady-state related calculations.

This short article proposes a novel continuous permutation method based on eigen-decomposition (CPED) to solve the correlation problem of wind power. The contributions of this article are summarized as follows.

1) An eigen-decomposition method is presented in the CPED to permute the sample matrix and remedy the defect of traditional Cholesky methods, which are unable to solve the correlation problem with the nonpositive definite correlation matrix.
2) A relative-error matrix is established to evaluate the deviation between wind speed samples and original wind speed data, and this deviation has not been discussed in previous studies of permutation and copula methods.
3) An iterative algorithm is proposed to achieve the continuous permutation and reduce the permutation deviation by adjusting the target correlation matrix. This algorithm is also applicable to other traditional permutation methods for improving accuracy.

II. PERMUTATION BASED ON EIGEN-DECOMPOSITION

The wind speeds are usually modeled as correlated random variables whose correlation information is described by the correlation matrix as

\[
    C_t = \begin{bmatrix}
        1 & \rho_{12} & \cdots & \rho_{1d} \\
        \rho_{21} & 1 & \cdots & \rho_{2d} \\
        \vdots & \vdots & \ddots & \vdots \\
        \rho_{d1} & \rho_{d2} & \cdots & 1
    \end{bmatrix}
\]

(1)

where \(\rho_{ij}\) is the Pearson correlation coefficient between the \(i\)th variable and the \(j\)th variable, and \(d\) is the number of random variables. The \(C_t\) is the target correlation matrix for permuting the samples. Suppose an \(n \times d\) sample matrix \(X_p\) is generated by \(n\) times sampling to these random variables

\[
    X_p = [\eta_1^T, \eta_2^T, \ldots, \eta_d^T]^T
\]

(2)

where samples of \(d\) random variables obtained by each sampling constitute one row of \(X_p\), and the vector \(\eta_k\) is correspondingly composed of samples of the \(k\)th random variable \((k = 1, 2, \ldots, d)\). The correlation matrix \(C_p\) of these samples can be calculated based on the sample matrix \(X_p\) [6]. Normally, \(C_p\) and \(C_t\) are different because specific sequences are used for the fast sampling [2]–[5]. Zhang et al. [6] proved that \(C_p\) can be changed to be closed to \(C_t\) by permuting the positions of entries in columns of \(X_p\). In this article, an order matrix \(R\) is built

Manuscript received 7 February 2021; revised 29 August 2021 and 17 December 2021; accepted 13 January 2022. Date of publication 11 February 2022. (Corresponding author: Nan Zhao.)

The authors are with the School of Electrical and Electronic Engineering, University College Dublin, 4 Dublin, Ireland (e-mail: quan.li@ucdconnect.ie; nan.zhao@ucd.ie).

Digital Object Identifier 10.1109/JSYST.2022.3144969

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
for the permutation. Compared with the traditional Cholesky method, the eigen-decomposition method is applicable to the situation that the $C_t$ is nondefinite matrix. The steps of the permutation are as follow.

1) Implement eigen-decomposition to $C_p$ and $C_t$

$$C_p = U_p A_p U_p^T, \quad C_t = U_t A_t U_t^T \tag{3}$$

where $A$ is the diagonal matrix of the eigenvalues of the corresponding correlation matrix and $U$ denotes the matrix composed of eigenvectors.

2) Build the order matrix $R$

$$R = X_p (U_t \sqrt{A_t} \times (U_p \sqrt{A_p})^{-1})^T \tag{4}$$

3) Permute the entries of $X_p$ according to the $R$. Suppose $x_{ak}$ represents the entries in the $a$th row and $k$th column of $X_p$, and $r_{bh}$ represents the entry in the $b$th row and $h$th column of $R$. Find rank($x_{ak}$) = rank($r_{bh}$) and then adjust the position of $x_{ak}$ in $k$th column from $a$th to $b$th row, where rank($\cdot$) represents the ranking of the entry in its column.

After the permutation in step 3), $X_p$ is updated to a new sample matrix $X_p'$. Noted, when $C_p$ or $C_t$ is a negative definite matrix, some entries of $R$ would contain imaginary part. In this case, the moduli of these entries are used for the sample matrix permutation, as the size of the imaginary part is much smaller than the size of the real part. Although this replacement may cause the deviation of the correlation, the deviation will be reduced to a given tolerance of the error through the continuous permutation.

Algorithm 1: The CPED.

1: Procedure CPED ($X_p, C_f$)
2: Set loop counter and flag: $k \leftarrow 0$, flag $\leftarrow 0$
3: Initial target correlation matrix: $C_t^0 \leftarrow C_t$
4: Initial sample matrix: $X_p^0 \leftarrow X_p$
5: Calculate correlation matrix of $X_p$: $C_p^0 \leftarrow \text{corrcor}_{e}(X_p)$
6: start loop: $k=0$
7: while flag $\neq 0$ do
8: Generate $E_m^p$ and $e_{sum}^k$ according to (5) and (6)
9: Solve (3) and obtain: $U_p^k, A_p^k, U_t^k, U_t^k$
10: Obtain the order matrix $R_k$ by (4)
11: Permute sample matrix: $X_p^{k+1} \leftarrow \text{permute}(X_p, R_k)$
12: # according to the order matrix $R_k$
13: $C_{p,k+1}^{k+1} = \text{corrcor}_{e}(X_p^{k+1})$
14: Update: $E_m^{k+1}$, $e_{sum}$ # by (5) and (6) with $C_{p,k}^{k+1}$ and $C_{t}^{0}$
15: $k \leftarrow k+1$
16: if $k > 30$ then
17: flag $\leftarrow 0$
18: else
19: for $i \leftarrow 1$ to $d$, $j \leftarrow 1$ to $d$
20: if $e_{ij}^{k+1} > \text{tolerance} \& e_{sum} \leq e_{sum}^k$ then
21: Update $C_{t,k}^{k+1} \leftarrow C_{t}^{k}$ by:
22: $\rho_{ij}^{k+1} \leftarrow \rho_{ij}^{k} \times (1 - \text{weight} \times e_{ij}^{k+1})$
23: # where $\rho_{ij}^{k}$ are the entries of $C_{t,k}^{k}$
24: end loop
25: return $X_p^k \leftarrow X_p^{k+1}$

In cases studies, weight $= 0.1$, tolerance $= 1 \times 10^{-3}$.

III. CONTINUOUS PERMUTATION

The deviation between the new sample matrix $X_p^*$ and the target correlation matrix $C_t$ is called permutation deviation, which can lead to error of the PPF calculation or static security assessment of the power system. A relative-error matrix $E_p$ is established for indicating the permutation deviation, whose entry $e_{ij}$ is defined as

$$e_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{\sigma_{ij} - \rho_{ij}}{\rho_{ij}} & \text{otherwise} \end{cases} \tag{5}$$

where $\sigma_{ij}$ and $\rho_{ij}$ are the entries at the $i$th row and $j$th column of $C_p$ and $C_t$, respectively. Also, the general error $e_{sum}$ is defined as

$$e_{sum} = \sum_{i,j} |e_{ij}| \times 100\% \tag{6}$$

The basic idea of the continuous permutation is to adjust $C_t$ according to the $e_{sum}$ and repeat the permutation with the adjusted $C_t$. Suppose the new sample matrix $X_p^*$ is generated by the $k$th permutation, namely $X_p^{k+1} = X_p^*$. By solving (5) and (6), $E_m^{k+1}$ and $e_{sum}^{k+1}$ can be also determined. Therefore, we can update the target correlation matrix to $C_t^{k+1}$ by adjusting its entries as

$$\rho_{ij}^{k+1} = \rho_{ij}^k \times (1 - \text{weight} \times e_{ij}^{k+1}) \tag{7}$$

where weight represents the step size. The detailed procedure of the continuous permutation is shown in Fig. 1 and Algorithm 1. The convergence condition of the continuous permutation is presented at the line 20 of Algorithm 1 where the procedure of the CPED is presented by the pseudocode. The corrcor$_e$ function in Algorithm 1 is to calculate the correlation matrix and permute ($X_p$, $R_k$) is to permutes $X_p$ according to order matrix $R_k$. The superscript $k$ of variables shown in Fig. 1 and Algorithm 1 denotes the number of iterations.

IV. CASE STUDIES

In order to study the performance of the CPED and its application, a 3-WFs IEEE 30-bus system [10] is used in this case running on a 2.6 GHz Intel Core i7 with 16 GB RAM. Three 30-MW (installed capacity) WFIs are integrated into the IEEE 30-bus system at bus-3, 5, and 7, respectively. The output of the active power $P(v)$ of each WF
can be expressed as

\[
P(v) = \begin{cases} 
0, & (0 < v < V_{ci} \text{ or } V_{co} < v) \\
\frac{v - V_{ci}}{V_{rs} - V_{ci}} \cdot P_N, & (V_{ci} \leq v \leq V_{rs}) \\
1, & (V_{rs} < v \leq V_{co})
\end{cases}
\]

where \(v\) is the wind speed and \(P_N = 30\) MW is the total installed capacity. Also, \(V_{ci}, V_{co},\) and \(V_{rs}\) denote the cut-in speed (\(V_{ci} = 3.5\) m/s), the cut-off speed (\(V_{co} = 25\) m/s) and the rated speed (\(V_{rs} = 9.5\) m/s), respectively. Two sets (Set-1 and Set-2) of wind speed samples are used as the original data of wind speed in WFs from the historical record in the northwest area of China [11], and each set of data contains 50,000 wind speed samples. The correlation information of the two sets of data is shown in Fig. 2.

Fig. 3 shows the general error \(e_{sum}\) of the sample matrix processed by different methods for data Set-1 (sampling size \(n = 400\)). After 14 iterations (14 times of permutations), the \(e_{sum}\) of the sample matrix processed by the CPED is converged to 0.27%, which is significantly lower than that of the Cholesky method and copula method (4.3% and 7.3%, respectively). It proves that the sample matrix permuted by the CPED retains the most correlation information of the original correlation matrix. Fig. 4 shows \(e_{sum}\) converges to 1.69% after 15 iterations of the CPED method with the data Set-2 whose correlation matrix is nonpositive definite. This error is larger than that in the data Set-1 case, which is mainly because some entries of the order matrix \(R\) for the permutation contain an imaginary part and the moduli of these entries are used for the sample matrix permutation. However, the \(e_{sum}\) of the sample matrix obtained by the CPED is still smaller than that of the copula method.

To show the performance of the proposed CPED applied to steady-state related calculations, the modified Latin hypercube sampling (LHS) [4] is combined with the CPED (LHS-CPED) and applied to calculate the PPF for wind power penetrated system, compared to the traditional LHS-Cholesky method. For data Set-2, where the correlation matrix is a nonpositive definite matrix, only the copula method [1] is used for comparison since the LHS-Cholesky method is not applicable for nonpositive definite matrix. The mean of the active power obtained by simple random sampling [4] with 50,000 times simulation is set as the reference and denoted as \(P_{b,s}\). The relative error of active power \(\varepsilon_{\mu}\) is used to describe the error of PPF results

\[
\varepsilon_{\mu} = \frac{1}{L} \sum_{b} \left| \frac{P_{b,a} - P_{b,s}}{P_{b,s}} \right| \times 100\%
\]

where \(L\) is the total number of branches of the grid, and \(P_{b,a}\) is the mean of the active power transmitted by the \(b\)th branch. Table I shows the \(\varepsilon_{sum}\) and \(\varepsilon_{\mu}\) of LHS-CPED are smaller than that of LHS-Cholesky and copula method in any size of sampling, reflecting the superiority of CPED in precision. Table II shows that the LHS-CPED method is more accurate than the copula method in the PPF calculation. In term of computation time for the PPF calculation, the LHS-CPED costs nearly the same time as the LHS-Cholesky method but less time than the copula method.

To further study the performance of the proposed method in a higher percentage wind penetration system, two more 10 MW WFs are added to the previous 3-WFs IEEE 30-bus system at bus-25 and 27, respectively, to make it a 5-WFs and 30-bus system. The wind speed samples used here are from Set-1. The correlation coefficient between new added WFs is 0.73, and the coefficient between new
added WF's with existing WF's is 0.42. In this case, the continuous permutation is combined with the Cholesky method (CP-CM) for the comparison. Fig. 5 shows $\varepsilon_{\text{sum}}^{P}$ of the sample matrix and the relative error $\varepsilon_{\mu}^{P}$ of the PPF results obtained by CP-CM and CPED (sample matrix is sampled by LHS with $n = 400$) with a different number of iterations. Also, as shown in Fig. 5, both $\varepsilon_{\text{sum}}^{P}$ and $\varepsilon_{\mu}^{P}$ are significantly improved by using the continuous permutation, especially when the iteration converges. It also indicates that continuous permutation can be applied for traditional permutation algorithms such as Cholesky method for better performance. In addition, $\varepsilon_{\text{sum}}^{P}$ and $\varepsilon_{\mu}^{P}$ of the CP-CM and CPED are significantly lower than that of the copula method.

Another case based on the IEEE 118-bus system [10] is carried out to investigate the performance of the proposed method in a large power system. Five WFs (30, 30, 30, 10, 10 MW) are integrated into the IEEE 118-bus system at the bus-4, 9, 38, 71, and, 83, respectively. The correlation matrix of the sample matrix is non-positive definite. Compared to the traditional Cholesky method and copula method, the CPED reduces the correlation deviation by more than 80%. It also shows that the reduction of the correlation deviation contributes to a lower error in the PPF calculation. Future work will be conducted to extend the application of the CPED to the static security analysis and optimal PPF calculation considering other kinds of uncertain energy sources, such as photovoltaic power generations.

V. CONCLUSION

This short paper proposes a novel method called CPED for wind power correlation analysis. The proposed method adopts the eigen-decomposition to realize the permutation process with a non-positive definite correlation matrix, which remedies the defect of the traditional Cholesky method. Besides, to reduce the correlation deviation, which is ignored in the existing studies, the continuous permutation process is proposed in this article. The case studies verify the effectiveness of the proposed CPED in the correlation analysis even the correlation matrix of the sample matrix is non-positive definite. Compared to the traditional Cholesky method and copula method, the CPED reduces the correlation deviation by more than 80%. It also shows that the reduction of the correlation deviation contributes to a lower error in the PPF calculation.

ACKNOWLEDGMENT

The authors would like to thank the China Scholarship Council (CSC) and University College Dublin (UCD) for provision of a UCD CSC Scholarship for Quan Li.

REFERENCES

[1] Z. Q. Xie, T. Y. Ji, M. S. Li, and Q. H. Wu, “Quasi-Monte Carlo based probabilistic optimal power flow considering the correlation of wind speeds using copula function,” *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2239–2247, Mar. 2018.

[2] S. Peng, J. Tang, and W. Li, “Probabilistic power flow for AC/VSC-MTDC hybrid grids considering rank correlation among diverse uncertainty sources,” *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 4035–4044, Sep. 2017.

[3] Y. Huang, S. Chen, Z. Chen, W. Hu, and Q. Huang, “Improved probabilistic load flow method based on D-vine copulas and latin hypercube sampling in distribution network with multiple wind generators,” *IET Gener., Transmiss. Distrib.*, vol. 14, pp. 893–899, 2020.

[4] Y. Chen, J. Wen, and S. Cheng, “Probabilistic load flow method based on nataf transformation and latin hypercube sampling,” *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 294–301, Apr. 2013.

[5] M. Hajian, W. D. Rosehart, and H. Zareipour, “Probabilistic power flow by Monte Carlo simulation with latin supercube sampling,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1550–1559, May 2013.

[6] S. Zhang, H. Cheng, L. Zhang, M. Bazargan, and L. Yao, “Probabilistic evaluation of available load supply capability for distribution system,” *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 3215–3225, Aug. 2013.

[7] P. Gao and D. Infield, “Wind turbine power curve modeling and monitoring with Gaussian process and SPRT,” *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 107–115, Jan. 2020.

[8] O. Z. Sehsalar, S. Galvani, and M. Farsadi, “New approach for the probabilistic power flow of distribution systems based on data clustering,” *IET Renewable Power Gener.*, vol. 13, pp. 2531–2540, 2019.

[9] J. Yu, W. Lin, S. Kamel, W. Li, and L. Zhang, “Equivalent model considering frequency characteristics and renewable uncertainties for probabilistic power flow,” *IET Gener., Transmiss. Distrib.*, vol. 12, pp. 5939–5948, 2018.

[10] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, “MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education,” *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[11] Hourly data from surface meteorological stations in China. [Online]. Available: http://101.200.76.197/en/?r=data/detail&dataCode=A.0012.0001