Research on Prediction Technology of UHV Transmission Line in Global Energy Internet

Lingyun Li 1,a,* Jingshuang Gan 2,b Xinju Guo 1 Yongjie Li 1 Xuyang Li 1 Tiejun Zhou 1
1 State Grid Henan Economic Research Institute Zhengzhou, China
2 Henan Electric Power Survey & Design Institute Zhengzhou, China
*a Corresponding author e-mail: 454174972@qq.com, b 1656402616@qq.com

Abstract. In response to the global energy crisis, the State Grid Corporation of China has put forward the idea of global energy Internet, which focuses on using ultra-high voltage (UHV) transmission technology for power delivery across continents. Reasonable prediction of UHV transmission line can effectively assist global energy Internet scheme comparison and project decision in the early stage. Based on mathematical model, this paper selects 11 influencing factors to obtain mathematical equation of UHV transmission line. It indicates that the mathematical model obtains good results and the deviation rate between the predicted value and the actual value is less than 10%. The prediction technology of UHV transmission line in global energy Internet can meet the requirements of scheme comparison and project decision in the early stage.

Keywords: Global energy Internet, UHV transmission line, prediction technology, mathematical model.

1. Introduction
Faced with the severe challenges of global energy security, environmental pollution and climate change, the State Grid Corporation of China (SGCC) has put forward the concept of global energy Internet, as in [1]. The development and utilization of Arctic wind energy and equatorial solar energy resources, and the use of ultra-high voltage (UHV) AC/DC transmission technology for power transmission across continents will be the development trend of the global power industry. It is estimated that by 2050, the electricity demand in Asia, Europe and North America will be $38 \times 10^{12}$ kWh, $9.5 \times 10^{12}$ kWh and $10.2 \times 10^{12}$ kWh respectively. The power supply from the Arctic and equatorial clean energy will be $3 \times 10^{12}$ kWh and $9 \times 10^{12}$ kWh. The large-scale transmission demand has put forward higher requirements and challenges to UHV transmission line, and it is very important to predict and control the cost of UHV transmission line project, as in [2].

Accurate determination of UHV transmission line project cost often requires technicians to extract engineering quantities from the designed schemes, apply engineering quantities and charges in accordance with the requirements and rules in the quota, and calculate the passage costs according to the relevant compensation documents of the project. This method generally requires technicians to determine the project schemes and utilize the quota accurately, and the whole process cycle is too long.
In the early stage of project decision-making, a technical method of UHV transmission line cost prediction is needed, which can not only reduce the work time, but also accurately predict the reasonable range of the project cost. At present, the domestic research on UHV transmission line project cost mainly focuses on cost influencing factors, cost control and management, and the establishment and research of cost prediction model are few, as in [3]. This paper attempts to establish a principal component regression model to predict the project cost with the 11 most important factors of UHV transmission line, which provides support for the scheme comparison and project decision of UHV transmission line in global energy Internet in earlier stage.

2. Cost Prediction Technology of UHV Transmission Line in Global Energy Internet

2.1. Principal Component Analysis Method

In the process of cost prediction of UHV transmission line, there are many influencing factors and the extent of the impact is different. Also, there may be collinearity among the influencing factors. In order to predict the cost accurately, it is necessary to eliminate the collinearity among the influencing factors using principal component analysis method. Indeed, principal component analysis extracts new variables by analyzing a large number of influence factors, and utilizes new variables to express the information of a large number of factors. The whole process not only eliminates multiple collinearity among factors, but also realizes dimension reduction.

In the application of principal component analysis, the sample data obtained is standardized, and the Bartlett test and KMO test are used to test whether the method is suitable for factor analysis and if it is applicable, the principal component analysis is carried out. In this paper, the principal component analysis of influencing factors is carried out by SPSS statistical software. The weighting coefficients of influencing factors are calculated by SPSS, and the factors whose eigenvalues are greater than 1 are extracted as principal component factors. Finally, the score of each factor is calculated through SPSS, and the factor score formula can be obtained as shown below in (1).

\[ F_i = b_{i1}A'_1 + b_{i2}A'_2 + \cdots + b_{in}A'_n, i = 1, 2, \cdots, n \]

In (1), \( A'_1, A'_2, \cdots, A'_n \) are normalized factors and \( b_{i1}, b_{i2}, \cdots, b_{in} \) are factor scores.

2.2. Principal Component Regression Model

Principal component regression model is to use principal component analysis method to eliminate multiple collinearity among influencing factors, get new principal component factor to express the information of original multiple factors, and then set up multiple linear regression models with new principal components as independent variables. Finally, the correlation and significance tests of the obtained principal component regression model are needed. The principal component regression model is shown in (2).

\[ Y^* = \beta_0 + \sum_{i=1}^{p} \beta_i F_i + \varepsilon \]
\[ \varepsilon \sim N(0, \sigma^2) \]

In (2), \( F_i (i = 1, 2, \cdots, p) \) are independent variables using principal component analysis, and \( Y^* \) is the standardized dependent variable, and \( \beta_i (i = 1, 2, \cdots, p) \) are regression coefficients and \( \varepsilon \) is the random error.
3. Cost Prediction Analysis of UHV Transmission Line in Global Energy Internet

3.1. Principal Component Analysis
In this paper, 11 factors influencing the cost of UHV transmission line are extracted, namely terrain, wire parameters, wind speed, icing, tower number, strain tower ratio, wire index, tower steel, foundation steel, soil and loop number, as in [4]. 50 samples are selected to get the corresponding data of influencing factors. After quantifying the qualitative data, the principal component regression model of the UHV transmission line cost prediction will be established by using 11 quantitative factors.

In this paper, the collinearity of 11 influencing factors is judged by SPSS software, and the multiple collinearity among factors is determined. After KMO test and Bartlett spherical test, the factors are judged to be suitable for factor analysis, and principal component analysis is carried out. Finally, three new principal component factors are extracted, as shown in Table I.

The score of factor is calculated by SPSS software, and the score expression of principal component factor is obtained, as shown in (3), (4) and (5).

\[
F_1 = 0.016A_1^* + 0.293A_2^* + 0.293A_3^* - 0.216A_4^* \\
+ 0.117A_5^* + 0.107A_6^* + 0.466A_7^* + 0.438A_8^* \\
+ 0.403A_9^* + 0.126A_{10}^* + 0.410A_{11}^* 
\]  

\[
F_2 = 0.571A_1^* - 0.005A_2^* - 0.057A_3^* + 0.084A_4^* \\
- 0.496A_5^* + 0.428A_6^* - 0.007A_7^* + 0.210A_8^* \\
+ 0.034A_9^* - 0.435A_{10}^* - 0.019A_{11}^* 
\]  

\[
F_3 = 0.072A_1^* - 0.369A_2^* - 0.448A_3^* + 0.309A_4^* \\
+ 0.312A_5^* + 0.412A_6^* - 0.134A_7^* + 0.150A_8^* \\
+ 0.337A_9^* + 0.364A_{10}^* + 0.094A_{11}^* 
\]

In (3), (4) and (5), \( A_1^*, A_2^*, \ldots, A_{11}^* \) are standard values of variables.

3.2. Establish the Regression Model
\( F_1, F_2 \) and \( F_3 \) are used as new independent variables and \( Y^* \) is used as dependent variable. Regression analysis is carried out and the results are shown in Table II-Table IV.

As can be seen from Table II, \( R^2=0.982 \) shows that the overall regression model is good, and \( F1, F2 \) and \( F3 \) account for 98.2% of the dependent variable. In Table III, \( p=0 < 0.05 \) indicates that the model has a strong significance at a significant level of 0.05, and the model is established.

In Table IV, the regression parameters of the model are given, and the P values of \( F1, F2 \) and \( F3 \) are all 0, indicating that each factor has a significant effect on the dependent variable \( Y^* \). The variance expansion factor VIF and tolerance are reciprocal, which is to determine the multicollinearity between variables. The larger the VIF of the independent variable, the less serious the tolerance is, the more serious
Table 1. Total variance of interpretation

| component | The initial eigenvalue | Extracting sum of squares |
|-----------|------------------------|---------------------------|
|           | Total Variance ( %)    | accumulation( %)          | Total Variance( %) | accumulation( %) |
| 1         | 4.195                  | 38.140                    | 4.195              | 38.140            |
| 2         | 2.492                  | 22.651                    | 2.492              | 22.651            |
| 3         | 1.755                  | 15.958                    | 1.755              | 15.958            |
| 4         | 0.801                  | 7.280                     | 0.801              | 7.280             |
| 5         | 0.699                  | 6.353                     | 0.699              | 6.353             |
| 6         | 0.411                  | 3.734                     | 0.411              | 3.734             |
| 7         | 0.354                  | 3.215                     | 0.354              | 3.215             |
| 8         | 0.167                  | 1.516                     | 0.167              | 1.516             |
| 9         | 0.098                  | 0.891                     | 0.098              | 0.891             |
| 10        | 0.020                  | 0.186                     | 0.020              | 0.186             |
| 11        | 0.008                  | 0.076                     | 0.008              | 0.076             |

Table 2. Model Summary

| Model | R       | R-squared | Adjusted R-squared | Error of the standard estimate | Durbin-Watson |
|-------|---------|-----------|--------------------|--------------------------------|---------------|
| 1     | 0.991   | 0.982     | 0.981              | 0.137                          | 1.415         |

Table 3. Regression Variance Analysis

| Model  | Sum of squares | df | Mean Square | F       | p.       |
|--------|----------------|----|-------------|---------|----------|
| Regression | 72.674         | 3  | 24.225      | 1296.781| .000a   |
| Residual error | 1.326         | 71 | .019        |         |         |
| Total   | 74.000        | 74 |             |         |         |

Table 4. Regression Variance Analysis

| Model  | Non-standardized coefficients | Standardized coefficients | t  | p  | Collinearity statistics |
|--------|-------------------------------|---------------------------|----|----|-------------------------|
|        | B                             | Standard error            |    |    | Tolerance | VIF |
| (Constant) |     -3.412E-16   | 0.016          | 0.000 | 1.000 |           |
| F1     | 0.424                        | 0.008          | 0.869 | 54.702 | 0.000 | 1.000 | 1.000 |
| F2     | 0.263                        | 0.010          | 0.415 | 26.117 | 0.000 | 1.000 | 1.000 |
| F3     | 0.176                        | 0.012          | 0.233 | 14.694 | 0.000 | 1.000 | 1.000 |

The collinearity problem is, and vice versa. It can be seen that the VIF values of the three factors are all less than 10, which indicates that there is no multicollinearity between the factors. This further shows that the principal component factor can solve the multicollinearity problem of multiple regression models.

Based on the above analysis, the regression model is verified by significance test, and there is no multicollinearity, and the regression coefficient is reasonable. The regression model is shown in (6).

\[ Y^* = 0.424F_1 + 0.263F_2 + 0.176F_3 \] (6)
The standardized variables are reverted to the original variables and the final regression model is obtained as shown in (7).

\[
Y = 17.420A_1 + 5.335A_2 + 2.141A_3 - 0.850A_4 \\
- 30.829A_5 + 3.472A_6 + 1.392A_7 + 0.669A_8 \\
+ 3.612A_9 + 0.718A_{10} + 42.234A_{11}
\]  

(7)

4. Verification of Cost Prediction of UHV Transmission Line

Five actual UHV transmission lines are randomly selected according to above models, and the predicted results are compared with the actual data to determine whether the model is feasible. The results are shown in Table V.

It can be seen from Table V that the deviation rate between the cost prediction and the actual cost of the five actual projects is less than 10%, which is consistent with the feasibility study of the line project. Therefore, the regression model established in this paper is reliable and can meet the project cost prediction of UHV transmission line in global energy Internet in the early stage.

| Sample | Estimated data (Ten thousand yuan/km) | actual data (Ten thousand yuan/km) | deviation rate (%) |
|--------|--------------------------------------|----------------------------------|-------------------|
| 1      | 630.21                               | 601.29                           | 4.81%             |
| 2      | 778.34                               | 759.07                           | 2.54%             |
| 3      | 524.89                               | 488.90                           | 7.36%             |
| 4      | 531.08                               | 528.70                           | 0.45%             |
| 5      | 710.34                               | 673.17                           | 5.52%             |

5. Conclusion

Global energy Internet proposed by SGCC which puts UHV power grid as the backbone network and delivering clean energy as the dominant, connects large-scale clean energy base and a variety of distributed power, and is a global energy configuration platform of wide service, strong energy configuration, high safety and reliability and green and low-carbon.

In this paper, the principal component regression model prediction technology is proposed to predict the cost of UHV transmission lines. It is proved that the prediction technology has good effect, and the deviation rate between the predicted value and the actual value is less than 10%, which can effectively support the scheme comparison and project decision of UHV transmission line in global energy Internet in the early stage. In the future work, it is necessary to continue to accumulate engineering data, improve the model constantly, and improve the accuracy of prediction, so as to realize the control of project cost in the early stage.

References

[1] Zhenya LIU, “Global energy Internet,” Beijing: China Electric Power Press, 2015
[2] Han GAO, Jichun LIU, “Analysis of transmission corridor planning in global energy Internet,” Sichuan Electric Power Technology, Vol. 40, pp. 15–20, 2017
[3] Huimin LIU, “Discussion on the effect of the project cost of UHV transmission line,” Shanxi Construction, vol. 27, pp. 221–222, 2013
[4] Wei Guo, Jing Wang, “Comparative analysis of cost and cost management between UHV and conventional transmission lines,” Value engineering, vol. 22, pp. 94–95, 2014
[5] Hua YE, Qianying MOU, Yutian LIU, “Enabling highly efficient spectral discretization-based eigen-analysis methods by Kronecker product,” IEEE Transaction on Power Systems, 2017, 32(5): 4148-4150