Application of principal component analysis method in comprehensive evaluation of soil corrosion

H Huang¹, Y Y Wu²,*, M F Peng², X H Ma¹, X Y Mao¹, J R Wu¹ and X Gao²
¹Electric Power Research Institute of Guizhou Power Grid Co., Ltd., Guiyang, China
²College of Electrical and Information Engineering, Hunan University, Changsha, China

Corresponding author and e-mail: Y Y Wu, 2584170656@qq.com

Abstract. The study of soil erosion in grounding grids is of great significance for ensuring the safe operation of power system. Through on-site testing of the influencing factors of soil erosion in the grounding grid of substations on 12 different regions of the country, the principal component analysis method was used to determine the correlation between soil corrosion factors and the corrosion rate of the buried test strips, and establish a comprehensive evaluation model for multiple indicators of soil corrosion in grounding grids. Research indicates, conductivity, Cl-, SO₄²⁻, K⁺ and Na⁺ are the key influencing factors of soil corrosion in the tested substation, and the corrosion score results that based on the proposed evaluation model are in good agreement with the corrosion rate of the buried test strips, so the comprehensive evaluation model of soil corrosivity indicators was successfully in the application of soil erosion studies and soil environmental assessment.

1. Preface

The grounding grids is the important facility in the power system to ensure the stable operation of the system and the safety of personnel[1]. At present, domestic power station and transformer substations commonly use flat steel or galvanized steel as the grounding mesh material, which have long been exposed to harsh underground operating environments and are subjected to soil erosion. The soil has a complex multiphase structure, which is composed of soil, water and air[2]. The factors that affect the soil's physical properties, chemical properties, electrochemical properties, and macroscopic environment all have an impact on soil erosion. Research shows[3-4], there are many factors affecting the corrosion of grounding grids in soil, including soil type, moisture content, pH, various anions and cations, total salt content, total voidage, total air, redox potential, resistivity, organic matter content. These influencing factors are related to each other and act together to accelerate corrosion of the grounding grids which are composed of steel.

When designing and researching the grounding grid, we must consider the corrosion situation of the soil where the grounding grid is located, study its soil erosion factors, and evaluate the corrosivity of the soil so as to predict the soil corrosion level of the grounding grid and take corresponding anti-corrosion measures in time. The corrosivity of soil is affected by many factors, and there is the cross-complex effect between each factor[5]. These factors’ comprehensive effect on soil erosion is not a simple linear superposition, if they are separately considered, it would be difficult to accurately
evaluate the corrosivity of soils by using single index evaluation or linear superposition evaluation of all factors. Therefore, it is necessary to study the relationship between soil erosion factors, and the principal component analysis method is an effective mean of de-correlation between factors. A comprehensive evaluation model of soil corrosivity was established by the principal component analysis method to construct a new comprehensive index that could reflect all the information of soil erosion and have no linearity with each other, so as to comprehensively evaluate the soil erosion, and select the key factors of soil erosion, and target anti-corrosion measures. In recent years, researchers at home and abroad tend to use various mathematical theory methods to evaluate the corrosive of soils. The principal component analysis method is a widely used and effective method in the study of soil erosion and soil environment evaluation[6]. In the field of soil erosion studies and soil environmental assessment, Zhu Zhiping[7] used the factor analysis method to evaluate the soil corrosivity of substation, Zhang Lianjin[8] used the factor analysis method to evaluate the soil corrosivity of Jiulongshan. Liu Xin[9] used the principal component analysis method to evaluate the soil quality of grassland in the Qinghai-Tibet Plateau. There are also scholars using the principal component analysis for the corresponding classification evaluation[10-11].

2. Principal component analysis

Principal component analysis (PCA) is an effective dimension reduction tool that commonly used in multivariate statistical analysis data[12].

2.1. Basic ideas

The principal component analysis method replaces the original variable by resolving the original variable with certain correlation into a new linearly independent integrated variable, and eliminates the correlation between the original variables. So we usually use the principal component analysis method to extract as few new variables that are irrelevant and can reflect all the original information of soil corrosive as possible.

2.2. Basic methods

2.2.1. Processing of original data

Since the influence of different soil corrosion factors on soil corrosion may be positively correlated or negatively correlated, the original datas need to be processed to meet the requirement of analytical calculations. The negative correlated influencing factors such as soil resistivity and pH value are in the form of reciprocal numbers to achieve the positive effect transformation.

2.2.2. Original data standardization

Since the corrosive factors of the soil have different dimensions, some of them have large differences in magnitude. In order to eliminate the influence of different dimensions and magnitudes, all original data must be standardized.

Assuming that the original variable data matrix is $X \in \mathbb{R}^{n \times p}$, and n represents the number of data sample, p represents the number of feature index, then standardizing the original data matrix to obtain a standardized matrix $Z \in \mathbb{R}^{n \times p}$, among them,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}$$  \hspace{1cm} (1)

$$S_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}$$  \hspace{1cm} (2)
2.2.3. Calculation of the covariance matrix
Since the standardized matrix has a mean of 0 and a variance of 1, the covariance matrix and the correlation matrix are identical. Therefore, we use the covariance matrix to reflect the degree of correlation between the various influencing factors, among them,

\[ v_{kj} = \frac{\sum_{i=1}^{n} (z_{ik} - \bar{z}_k)(z_{ij} - \bar{z}_j)}{\sqrt{\sum_{i=1}^{n} (z_{ik} - \bar{z}_k)^2} \sqrt{\sum_{i=1}^{n} (z_{ij} - \bar{z}_j)^2}} \]  

(3)

2.2.4. Calculation of the eigenvalue and eigenvector of the covariance matrix
Setting the eigenvalues of the covariance matrix is \( \lambda_i \) \( (i = 1, 2, ..., p) \), and the eigenvector matrix is \( U \in R^{p \times p} \), so you can get,

\[ U^{-1}VVU = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}_{p \times p} \]  

(4)

Combining formula (3), (4), we can obtain the eigenvalues and the corresponding standardized orthogonal eigenvector matrix, and the eigenvalues can reflect the influence degree of the corresponding principal component features.

2.2.5. Determine the number of principal elements
In fact, we do not need all \( p \) principal elements. Generally, it is only necessary to select a reasonable part to reflect almost information of \( p \) principal elements, so we usually select \( m \) \( (m < p) \) principal components. The first \( m \) principal components can be selected by calculating the variance contribution rate \( w_j \), and the cumulative variance contribution rate \( \alpha_k \) is guaranteed to be not less than 85%.

\[ w_j = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda_i} \quad (j = 1, 2, \cdots, p) \]  

(5)

\[ \alpha_k = \sum_{i=1}^{k} \lambda_i \bigg/ \sum_{i=1}^{p} \lambda_i \quad (k = 1, 2, \cdots, p) \]  

(6)

2.2.6. Calculate the principal component comprehensive sample matrix
Setting \( T \) is a matrix that formed by standardized orthogonal eigenvectors corresponding to the selected eigenvalues of the first \( m \) principal components. Since the principal component comprehensive sample matrix is a projection of the standardized matrix \( Z \) in the direction of the first \( m \) principal components, so the matrix is expressed as:

\[ F = ZT \]  

(7)

3. Mathematical models for comprehensive evaluation of multiple indicators
At present, there are few studies on the evaluation methods of soil corrosion in grounding grids. Generally, the soil corrosion is evaluated by the single indicator, such as soil resistivity, but the factors that affect soil corrosion are numerous and complex, and they are closely related to each other. Therefore, the single indicator evaluation method does not fully reflect the extent of soil corrosion. In order to avoid the blindness of the evaluation and increase the accuracy of the evaluation, it is
necessary to study the comprehensive evaluation model of multiple indicators to accurately assess the soil corrosion status of the grounding grids. Firstly, the principal component analysis is used to correlate all the influencing factors, and the principal component that can reflect all the characteristic information of the soil corrosivity is obtained. Then calculating the weight scores corresponding to each principal component, and accumulating them to obtain the corrosion score which could characterize the extent of soil corrosion.

Therefore, establishing the comprehensive evaluation model for multiple indicators:

\[
f(t) = \alpha_1 f_1(t) + \alpha_2 f_2(t) + \cdots + \alpha_m f_m(t) + \cdots + \alpha_p f_p(t)
\]

\[
\alpha_1 + \alpha_2 + \cdots + \alpha_m + \cdots + \alpha_p = 1
\]

\[
\alpha_i \geq 0 \quad i = 1, 2, \ldots, p
\]

\(f(t)\) is the corrosion scores, \(f_i(t)\) represents the new principle component obtained by principal component analysis, \(\alpha_i\) is the weight score (or the principal component contribution rate). \(m\) refers to the first \(m\) principal components that are selected from \(p\) principal component.

By substituting the new principal component \(f_i(t)\) and its corresponding weight score \(\alpha_i\) of each sample into formula (1), the corrosion score of each sample can be obtained to represent the degree of the soil corrosion.

4. Application

4.1. Test point selection

In order to verify the feasibility and accuracy of the proposed model for comprehensive evaluation of multiple indicators for soil corrosion of grounding grids, we selected 25 sets of data as test sites from 12 substations with alkaline soils in the country, followed by: Nanchong Station 1; Jinan Station 2, 3, 4, 5; Xi’an Station 6, 7, 8, 9; Changxindian Station 10, 11, 12; Cold Lake Station 13; Daqing Central Station 14, 15; Changdu Central Station 16, 17; Dagang Central Station 18; Dunhuang Station 19, 20; Xiningjiao Central Station 21; Yumen Station 22, 23; Wuzhou Station 24, 25.

4.2. Selection of factors affecting soil corrosion

The 12 kinds of physical and chemical factors that have important effects on soil erosion, such as pH value, water content, conductivity, organic matter, total nitrogen content, and HCO\(^{-}\), Cl\(^{-}\), SO\(_4^{2-}\), Ca\(^{2+}\), Mg\(^{2+}\), K\(^{+}\), Na\(^{+}\) were collected as the index data for evaluating soil corrosion at the soil sites of the 12 substation grounding grids. Meanwhile, test strips were placed at the test points of each substation and the corrosion rate of the test strips was measured by using the weight loss method.

4.3. Results and analysis

Based on the soil corrosion factors of 12 substation grounding grids that with alkaline soil in the country, the principal component analysis method was used to determine the correlation between the soil corrosion factors of each substation and the corrosion rate of the test strips, and the principal elements were extracted. Finally, constructing the comprehensive evaluation model of soil corrosion.

The correlation shows that each corrosion factor has a certain impact on the corrosion rate, and the most influential factor is conductivity, followed by Na\(^{+}\), K\(^{+}\), Cl\(^{-}\). Water content, conductivity, and the anions, positive ions are positively correlated with the corrosion rate, while pH, organic matter, total nitrogen content and corrosion rate are negatively correlated. Meanwhile there is a strong correlation between soil corrosion factors.

The principal component analysis method eliminates the correlation between various corrosion factors and extracts the four most important principal component F1, F2, F3 and F4, which their cumulative variance contribution rate reaches 86.08%, that is, the influence of all soil corrosion factors in 12 tests area on soil corrosion can be reflected 86.08% of information by these four principal components. The eigenvalues, variance contribution rates and cumulative variance contribution rates
of these four principal components are shown in table 1. The variance contribution rate of F1 factor accounts for 49.30% of the total variance contribution rate, which is much higher than other factors. Therefore, F1 factor has a decisive significance to the composition of key soil corrosion factors.

| principal component | eigenvalues | variance contribution rate | cumulative variance contribution rate |
|---------------------|-------------|-----------------------------|---------------------------------------|
| F1                  | 5.9162      | 0.4930                      | 0.4930                                |
| F2                  | 2.1118      | 0.1760                      | 0.6690                                |
| F3                  | 1.5602      | 0.1300                      | 0.7990                                |
| F4                  | 0.7411      | 0.0618                      | 0.8608                                |

At the same time, the load distribution of principal component in soil corrosion factors is analyzed, F1 contains five factors such as conductivity, Cl\(^-\), SO\(_4^{2-}\), K\(^+\), Na\(^+\), which have highly positive correlations each other and commonly affect the soil corrosion. And F1 has a higher positive load on the concentration of these five factors. Therefore, the conductivity, Cl\(^-\), SO\(_4^{2-}\), K\(^+\), Na\(^+\) are the key affecting factors of the soil corrosion in the grounding grids of the 12 substations that tested.

Studies have shown that with the concentration of soil conductivity, Cl\(^-\), SO\(_4^{2-}\), K\(^+\) and Na\(^+\) increases, the corrosion of metal materials in grounding grids can be significantly accelerated. Soil conductivity is the main indicator of the ability of soil conductive. In the alkaline conditions, the higher the soil conductivity, the stronger the corrosion. Cl\(^-\) is the most influential anion in soil corrosion, which it can destroy the passivation protective film of metal or change the solubility of corrosion products through its adsorption and penetration, thereby promote the corrosion of the anode process. And it is also decisive for the occurrence and development of local corrosion. As a result, the higher the Cl\(^-\) concentration in the soil, the stronger the corrosiveness of the soil. SO\(_4^{2-}\) has a certain influence on the soil pH value, which it can enhance the acidity of the soil, lower the pH value, and indirectly promote the corrosion of the soil. K\(^+\) and Na\(^+\) have a strong correlation with conductivity, have no direct effect on soil erosion, but they can promote soil conductivity and accelerate soil erosion. In summary, the corrosion of the key factors such as soil conductivity, Cl\(^-\), SO\(_4^{2-}\), K\(^+\), Na\(^+\) in soil is consistent with the correlation obtained by the principal component analysis method.

In order to verify the feasibility of the comprehensive evaluation model of multiple indicators that established by principal component analysis in soil corrosion, we used Matlab software to simulate the variation rule of corrosion scores of different test points in each substations, as well as the corrosion rate of buried test strips and the conductivity that have the greatest effect on the soil corrosion, as shown in figure 1.

From figure 1, it can be seen that the degree of coincidence between the variation rule of the corrosion scores of different test points in each substations and the corrosion rate of the test strips reach 95.24%, but the variation rule of conductivity and corrosion rate is only 57.14%. The above shows that the corrosion score obtained by the proposed model is in good agreement with the variation in the corrosion rate of the test strips, while the evaluation of the single factors conductivity results in a large error. Therefore, the application of the comprehensive evaluation model for multiple indicators is successful in soil corrosion studies of grounding grids.
5. Conclusions
Through on-site testing of the factors affecting the soil corrosion of grounding grids in 12 different substations in the country, the principal component analysis method is used to analyze the factors of soil corrosion. Results indicate that there is a strong correlation between soil corrosion factors, and these factors all have a certain impact on the rate of soil corrosion. Among them, conductivity, Cl⁻, SO₄²⁻, K⁺, and Na⁺ are the key corrosive factors in the soil. At the same time, the comprehensive evaluation model of multiple indicators of soil corrosion in grounding grids was established. The simulation results of Matlab software verify that the corrosion scores obtained by this model are in good agreement with the corrosion rate of on-site buried test, which proves that the application of the model in the study of soil corrosion and the soil environmental assessment is completely feasible.

The study on the corrosivity evaluation of grounding grid soil has only just begun. In fact, there is a big difference in magnitude in the corrosion scores obtained by the comprehensive evaluation model for multiple indicators and the actual corrosion rates. Therefore, the corrosion score obtained by the evaluation model cannot be directly used to evaluate the degree of soil corrosion, and the degree of corrosion can only be judged by comparison between each corrosion scores. At present, the country has not yet established a comprehensive set of corrosion evaluation standards. Therefore, it is necessary to strengthen research, and put forward to a set of standardized and fully applicable comprehensive evaluation model, so as to accurately determine the corrosion of the soil in grounding grids.

Acknowledgment
This work is supported by the National Natural Science Foundation of China (No. 61472128, 61173108).

References
[1] Zhang J H 2016 The Study of Grounding Grid Corrosion Electrochemical Measurement Method (North China Chongqing University)
[2] Li Y Q, Du C W, Feng H and Zhao Y 2010 *Soil Corrosion of Grounding Electrode (Environmental Testing)* p 15-18

[3] Li S F, Chen Z Z and Peng M F 2003 *Corrosion analysis and anti-corrosion technology to earthing network system (Inner Mongolia Electric Power)* chapter 21 p 9-12

[4] Shao J R and Le Y T 2006 *Study on corrosion and protection of AC system grounding net (Electrical Equipment)* p 70-72

[5] Wu D, Fu L and Qiu C H 2014 *Corrosion Evaluation and Corrosion Protection Strategy of Grounding Grid (Pioneering With Science & Technology Monthly)* p 171-172+181

[6] Zhang G Y, Liu S M, Chen C Y, Xie D and Shun H Z 2007 *Soil factors affecting corrosion of grounding grids in CangZhou, HeBei Province (Acta Pedologica Sinica)* p 1036-1042

[7] Zhu Z P, Wang L J, Pei F, Yin C H, Wu F Y and Tian X 2014 *Evaluation of Soil Corrosivity for Ground Grid of Substations in an Area by Factor Analysis (Journal of Chinese Society for Corrosion and Protection)* p 147-152

[8] Zhang L J, Lai G H, Kong Y and Sun C Z 2016 *Evaluation of Soil Quality in Beijing Jiulong Mountain Based on Factor Analysis Method (Journal of Northwest Forestry University)* p 7-14

[9] Liu X, Wang Y B, Lv M X, Sun Y, Yang W J and Zhao J P 2018 *Soil quality assessment of alpine grassland in permafrost regions of Tibetan Plateau based on principal component analysis (Journal of Glaciology and Geocryology)* p 469-479

[10] Liu J X and Tang Y H 2018 *Evaluation of Soil Fertility Grade for Old Trees Based on Principal Component Analysis (Anhui Agric Sci)* p 104-106

[11] Rong Y, Liu Z B, Feng J Y and Li L 2007 *Application of principal components analysis to analyze of sewage irrigation region (Journal of Liaoning Technical University)* p 252-254

[12] Yang H, Zhou W, Wen H R, Zhao H W and Zhang W S 2007 *The Application of Principal Component Analysis (PCA) Method in Evaluation of Soil Corrosivity (OGST)* p 43-46+62+1