

## Electric Field Strength Calculation of Regression Model Based on Principal Component Analysis

Zhuoshi Li<sup>1, a</sup>, Zenghui Wang<sup>2, b \*</sup>, Tingting Zhang<sup>3, c</sup>, and Xiaoqi Ding<sup>4, d</sup>

<sup>1,3,4</sup> Jilin Agricultural University, Changchun, Jilin, 130018, China

<sup>2</sup> College of Humanities & Sciences of Northeast Normal University, Changchun, Jilin, 130017, China

<sup>a</sup>leezs643@sina.com, <sup>b</sup>wzh195693@126.com, <sup>c</sup>499360221@qq.com, <sup>d</sup>leezs036@foxmail.com

**Keywords:** transmission power; frequency electric field; principal components analysis and multiple regression analysis

**Abstract.** In light of the transmission line electromagnetic environment of the increasingly prominent problems, this paper mainly studies the related factors of the transmission line frequency electric and magnetic field, using the method of principal component analysis to extract the main factors influencing the intensity of electric field; Using the variable and the correlation between variables, the electric field intensity achieved by multiple regression analysis method of regression model, giving help to the design of high-voltage wire transmission line.

### Introduction

In recent years, the problem of electromagnetic environment caused by power transmission and transformation project is more and more prominent, there have been all over the country residents petition, block construction, power supply departments to sit quietly, and so on and so forth. Investigate its reason mainly has the following several aspects: (1) the lack of mandatory national standards, high voltage electric power transmission and transformation project. The national standard of power frequency electromagnetic field problems of research work is under way (2) the power construction planning and urban construction planning are not synchronized. Power construction built from planning to take a long time, when an electric power construction project for examination and approval of planning, design, finished, in the process of construction, due to the change of the urban construction planning, some originally built in the city periphery also began to develop around the high tension line and so on electric power facilities, leading to new problems (3) because people don't understand the characteristic of electromagnetic field or electromagnetic waves, especially for power frequency electric power transmission and transformation projects, the misunderstanding of the magnetic field, the high voltage transmission line and electromagnetic radiation even linked with radiation; (4) the substandard construction of living community.

### The Basic Idea of Principal Component Analysis Method

Dimension principal component analysis (pca) is a kind of method, namely on the premise of information loss rarely, multiple indicators into several comprehensive index of multivariate statistical methods. Conversion generated after the comprehensive index of said main ingredients, it is formed by a linear combination of the original variables, and unrelated between each main component, so the principal component analysis, compared to the original variable has more superior performance. It solves the multicollinearity problem between multiple variables. Complex problems need to introduce multiple variables, you can consider only a few principal components to replace, such not only won't lose too much information, but also more able to seize the main contradiction, revealing the internal variables, the regularity between the things made the research question is simple, the efficiency is improved.

## The Basic Steps of Principal Component Analysis

Electric field intensity is describe the electric field of a physical quantity. Some observation points in the field of electric field intensity  $E$ , is equal to the static test at that point charge  $q$  'force  $F$  and the power of  $q$  ratio. Test charge  $q$  'value should be small enough, don't change it in place of the electric field. In this way, the electric field intensity is positive test charge per unit force. There are many factors that can influence of electric field intensity, such as: circuit current, circuit voltage, the number of division, the environment temperature, environment humidity, altitude, atmospheric pressure, the line height, the power frequency electric field intensity of measuring point, etc. On the influence of electric field intensity each are not identical.

Line current: when choosing wires, switch and fuse, will involve the issue of account of load current.

Line voltage: for three phase four wire system of power grid, the three phase line voltage between any two root said line voltage, any one phase line and between the zero line voltage phase voltage, three-phase voltage phase difference of 120 degrees, line voltage are the two phase voltage vector and phase, the size of the line voltage and phase voltage relations are: line voltage = phase voltage of the square root of 3 times. Regarding the mains, phase voltage 220 v.

Division number: reduce the line reactance, reduce conductor surface electric field intensity, and other general purpose use split conductor is 220kv for 2 division , 500kv for 4 division, 750kv for 6 northwest power grid division, 1000kv for 8 division.

Ambient temperature: power frequency electric field with more attention paid to the influence on human living environment, which temperature changes lead to the change of power frequency electric field measurement values.

Environmental humidity: power frequency electric field with more attention paid to the influence on human living environment, the humidity changes lead to the change of the power frequency electric field measurement value, discusses the humidity increases, the main reason of the change of the measuring instrument performance and test verification.

Altitude: altitude also calls the absolute height, is a place with the sea level is poor, usually do to mean sea level. standard represent the ground vertical distance of a place above sea level.

The atmospheric pressure: the atmospheric pressure is the earth's surface is covered with a thick layer of the atmosphere of air. The objects in the atmosphere, to be by the pressure of the collision of the air molecules. Also can be thought of, the atmospheric pressure is objects in the atmosphere by the atmosphere of its own gravity pressure acting on the object.

Line height: electric field intensity with the increase of wire to ground distance, decreased significantly at the beginning, after decreasing degree gradually slow. Is effective to increase the line height to reduce the field strength, the price is need to increase investment.

The divided interval: times conductor with the increase of operating voltage division spacing influence on maximum nominal field strength increases slightly.

## The Established Model

### The Selection of Indicators

$X_1$  the line current,  $X_2$  the environment humidity,  $X_3$  altitude,  $X_4$  environment temperature  $X_5$  atmospheric pressure,  $X_6$  line height,  $X_7$  division number,  $X_8$  point power frequency electric field intensity,  $X_9$  line voltage,  $X_{10}$  divided interval.

### Data Processing

Use the principal component analysis, it is necessary to determine whether there is a strong correlation between variables. If the variable correlation between weak, using principal component analysis can't have very good dimension reduction effect. So first to analyze the correlation between variables correlation matrix between the selected variables.

Most of the coefficient of correlation coefficient is greater than 0.3, shows a strong linear relationship between variables, suitable for family composition analysis.

Table 1 The common factor variance

	initial	extract
VAR00001	1.000	0.756
VAR00002	1.000	0.629
VAR00003	1.000	0.902
VAR00004	1.000	0.810
VAR00005	1.000	0.929
VAR00006	1.000	0.583
VAR00007	1.000	0.927
VAR00008	1.000	0.702
VAR00009	1.000	0.877
VAR00010	1.000	0.868

This Analysis is to use Factor Analysis module is the default information extraction method, principal component Analysis.Can see in addition to the degree of information loss is bigger, the principal component contains almost all of the original variables most of the information.

Total variance analysis table shows the main component of the total variance explained the original variables.In this case, see reserve four principal component is advisable.The front three principal components, the variance and the proportion of the total variance is 79.836%, we will select the  $X_1$  as the first principal component, the second principal component  $X_2$ ,  $X_3$ , is the third principal component,  $X_5$  for the fourth principal component.

Namely after the extraction of principal component analysis method, we know that the line current, environment humidity, altitude, environment temperature, they are in power frequency electric field intensity as the main influencing factors.

### The Multiple Regression Analysis

Multiple linear regression model, if there are two independent variables, and the same are in linear relationship between the dependent variable, at this time as the binary linear regression model, regression model is described:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e \quad (1)$$

Set  $y$  as the dependent variable,  $x_1, x_2, \dots, x_j$  as the independent variable, and when to the linear relationship between independent variable and dependent variable, the multivariate linear regression model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j + e \quad (2)$$

Among them,  $\beta_0$  for the constant term,  $\beta_1, \beta_2, \dots, \beta_j$  For the regression coefficients in table 2 and table 3.

Table 2 Anova

model	Square	df	mean square	F
regression	141.3327	4	353.3319	9.303
residuals	607.7024	160	379.814	
total	749.0351	164		

Table 3 Coefficient

model	non-standardized		standard		t	Sig.
	coefficients		coefficients			
	B	standard error				
constant	-1247.060	521.58			-2.391	0.018
VAR00001	-0.100	0.448	-0.018		-0.222	0.825
VAR00002	557.375	633.89	0.073		0.879	0.381
VAR00003	6.965	1.215	0.556		5.734	0.000
VAR00005	30.918	12.261	0.258		2.522	0.013

Regression analysis is to identify the two or more variables quantitative relationship interdependence between a statistical analysis method. Use is very extensive, regression analysis, according to how much involved in the independent variables are divided into regression and multiple regression analysis; According to the amount of the dependent variable, can be divided into a yuan regression analysis and multiple regression analysis; According to the relationship between the dependent and independent variables and types, can be divided into linear regression analysis and nonlinear regression analysis. If in regression analysis, including only one independent variable and a dependent variable, and the relationship of the two factors can be used in a straight line to approximate, the regression analysis, called monadic linear regression analysis. If included in the regression analysis of two or more independent variables, and is a linear relationship. Between the dependent variable and independent variables, it is called a multiple linear regression analysis. This paper use the multiple regression analysis to data to establish a regression model:

$$Y = -0.018X_1 + 0.73X_2 + 0.556X_3 + 0.258X_5 \quad (3)$$

By above, The power frequency electric field strength and the line current, environment humidity, altitude, closely related to environmental temperature, this paper combines the principal component analysis in the application of the electric field and the combination of regression analysis, we adopt a few less instead of more variables, and thus simplify the complexity of the problem, we apply this method in the field, the problem of the electric field was studied, and combined with the regression analysis to the problem of electric field is forecasted, and achieve expected effect.

## Acknowledgements

The Youth Foundation of Jilin Agricultural University (201328)  
 Jilin Province Innovation and Entrepreneurship Students Training Program Project  
 Jilin Agricultural University Undergraduate Science and Technology Innovation Fund Project

## References

[1]Guo.Z.G, Multiple regression analysis method in evaluating the application of the joint-stock company profitability, Journal of mathematical statistics and management, (1995)21-23.

- [2] Maruvada P, An experimental study of residential fields in the vicinity of transmission lines, IEEE Trans on power delivery. (1998)56-60.
- [3] Yang.X.D, Engineering electromagnetic field numerical analysis and synthesis, Mechanical industry publishing., Beijing, 2008.
- [4] Zhang.Q.C, Mathematical model of high-voltage overhead power frequency electric field near the line, Electric Power Environmental Protection, 4 (2000)14-17.