

PAPER • OPEN ACCESS

Analysis of Key Environmental Impact Factors Based on Fuzzy Clustering and Grey T's Correlation Degree Method

To cite this article: Gaofei Zhao *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **472** 012079

View the [article online](#) for updates and enhancements.

Analysis of Key Environmental Impact Factors Based on Fuzzy Clustering and Grey T' s Correlation Degree Method

Gaofei Zhao¹, Yujie Huo¹, Zhaofeng Zhang¹ and Zhaoyuan Peng²

¹Beijing HAZT Testing Technology Co., Ltd. No. 9 Shun Xing Road, Shunyi District, Beijing

²Beijing Suplet Power Co., Ltd.
Email: zhaogaofei@hazt.cn

Abstract. This paper expounds the basic principle of fuzzy clustering and grey T' s correlation degree method. It takes a power module used in communication device and worked outfield as an example to analysis the impacts of natural environment factors on product, that based on failure rate. Firstly, fuzzy clustering method is used to classify the environmental factors into four categories. Secondly, grey T' s correlation degree method is used to compare and rank the degree of influence of the four types of factors on the failure rate of power module. Finally, the analysis method and the key environmental factors affect the failure of the power module are given. The analysis of key influencing factors lays a foundation for the analysis of failure causes and mechanism of products, and provides a basis for reliability test design.

1. Introduction

In recent years, with the wide application of electronic products, the reliability of electronic products has been increasing, especially those used in outdoor environments. Communication equipment has been distributed throughout the country and a power module must be with them. Our territory is vast and the natural environment is complex and changeable. Therefore, the power supply module of communication equipment is subjected to various harsh natural climatic factors, and the reliability of the product is threatened. We can't intuitively determine which factors have the same or similar effects on product failure and which factors play a major role in the process of product failure because of the complex and variable environmental impact factors. This brings difficulties to the analysis of fault mechanism, the design of acceleration test and the design of product reliability.

Cluster analysis can classify complex environmental factors and classify similar factors into the same category. We choose representative factors to characterize such environmental factors and achieve the purpose of reducing species of the influencing factors [1]. Jia Zhijun used principal component analysis to evaluate the climatic factors of 13 atmospheric corrosion sites in China. He divided the climatic factors of the 13 sites into 6 categories [2]. Zhang Lunwu classified the environment of 8 atmospheric corrosion test stations by fuzzy clustering based on the environmental factors of the atmospheric corrosion test stations in China for more than ten years. The correlation between 17 environmental factors affecting atmospheric corrosion was comprehensively evaluated [3]. But the final result of cluster analysis is to classify environmental factors, which can't reflect the impact of environmental factors on product failure. Through clustering analysis, there are still many kinds of stress after the reduction of species. It is difficult to identify and analysis the key causes of product failure. We need to further analysis and calculate the key factors that affect product failure.

Grey correlation analysis can describe the correlation between various variables effectively and has a good measure for various nonlinear correlation relations [4], [5]. The grey system theory proposes



the concept of grey correlation analysis for each subsystem, and intends to seek numerical relationships between subsystems or factors in the system through certain methods. Therefore, grey correlation is an effective means to measure the closeness of two variable factors [6].

Based on the weather factors and failure rate data of the service area of power module within a communication apparatus, this paper uses fuzzy clustering method to cluster the environmental influencing factors. Then the key factors affecting the fault are quantitatively extracted by means of grey correlation analysis, which lays a foundation for the analysis of fault mechanism and reliability test of design.

2. Fuzzy Clustering Analysis of Environmental Impact Factors

2.1. Fuzzy Cluster Analysis Method

The results of cluster analysis reflect the potential structure, characteristics and similarity of the raw data [7]. In general, cluster analysis can be divided into two categories. They're absolute clustering and fuzzy clustering. Absolute cluster analysis means that there is only one class identifier for a data object while data objects belong to multiple classes at the same time with different probabilities in fuzzy clustering. The specific processes are as follows.

- List the original data matrix

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{bmatrix} \quad (1)$$

X_{ij} is the average number of i th environmental factors in the j th time period. N is the number of environmental factors. M is the number of time periods, so the matrix $X=[X_{ij}]_{n \times m}$ is obtained in the formula.

- Transform the value to a range

$$X_{ij}^1 = \frac{X_{ij} - \wedge_{i=1}^n (X_{ij})}{\vee_{i=1}^n (X_{ij}) - \wedge_{i=1}^n (X_{ij})} \quad (i=1,2, \dots; n) \quad (2)$$

The denominator is the difference between the maximum and the minimum of the j column. Obviously

$$X_{ij}^1 \in [0, 1], \text{ so } X^1 = [X_{ij}^1]_{n \times m}.$$

- Seeking similarity relation matrix

The similarity relation matrix is obtained by the maximum minimum method as follow.

$$r_{ij} = \frac{\sum_{k=1}^m (X_{ik}^1 \wedge X_{jk}^1)}{\sum_{k=1}^m (X_{ik}^1 \vee X_{jk}^1)} \quad (3)$$

r_{ij} is the similarity coefficient between the environmental factor I and the environmental factor J . m is the number of environmental factors. X_{ik}^1 is the value of the I environmental factor in the K period.

X_{jk}^1 is the value of the j th environment factor in the k th period. The formula $X_{ik}^1 \wedge X_{jk}^1$ represents the minimum value between the X_{ik}^1 and X_{jk}^1 . $X_{ik}^1 \vee X_{jk}^1$ indicates the maximum value between the two.

According to the formula (3), we can get similar matrix $R = [r_{ij}]_{n \times n}$ by processing related data.

- Equivalence relation matrix

Using the transfer closure method to find the equivalent relation matrix

$$t(R) = \bigcup_{i=1}^{\infty} R^i \quad (4)$$

- Drawing dynamic cluster diagram

Draw dynamic cluster diagram and do cluster analysis by selecting appropriate similarity coefficient lambda value and using the equivalence relation matrix.

2.2. Data Collection of Environmental Factors

The collection of environmental factor data refers to the study of natural environment and climate factors for the areas in which the subjects are servicing. If a certain type of power module is mainly served in the A area, we conduct data analysis and collation of the climatic factors that may affect the failure rate of A. As shown in Table 1, all climatic factors in the table mean the average of this month.

Table 1. Climate Statistics of the Area A in 2017

2017	Temperature	humidity	PM2.5	radiation intensity	wind speed	Precipitation	sunshine time
January	3.32	78.13	114.29	2.81	2.58	0.25	6.8
February	7.64	64.39	69.86	3.69	2.64	2.03	8.59
March	14.13	67.26	62.68	4.80	2.61	7.87	8.45
April	23.70	72.33	53.27	5.86	2.90	0	8.86
May	29.77	74.10	59.42	6.36	3.61	28.95	9.09
June	30.97	81.83	42.47	6.05	2.60	80.78	7.48
July	31.61	95.03	52.06	5.22	2.00	63.51	4.95
August	30.13	97.71	38.06	4.93	2.06	199.9	7.05
September	27.87	95.23	57.40	4.57	2.2	3.05	6.72
October	17.23	97.45	56.94	3.73	2.32	32.77	3.99
November	9.57	80.23	45.43	2.80	2.83	0.00	5.15
December	4.48	72.48	43.65	2.41	2.58	0.00	5.32

2.3. Cluster Analysis of Environmental Factors

The original data matrix of natural environment factors for area A in 2017 is as follows.

$$X = \begin{bmatrix} 3.32 & 7.64 & 14.13 & 23.7 & 29.77 & 30.97 & 31.61 & 30.13 & 27.87 & 17.23 & 9.57 & 4.48 \\ 78.13 & 64.39 & 67.26 & 72.33 & 74.1 & 81.83 & 95.03 & 97.71 & 95.23 & 97.45 & 80.23 & 72.48 \\ 114.29 & 69.86 & 62.68 & 53.27 & 59.42 & 42.47 & 52.06 & 38.06 & 57.4 & 56.94 & 45.43 & 43.65 \\ 2.81 & 3.69 & 4.8 & 5.86 & 6.36 & 6.05 & 5.22 & 4.93 & 4.57 & 3.73 & 2.8 & 2.41 \\ 2.58 & 2.64 & 2.61 & 2.9 & 3.61 & 2.6 & 2 & 2.06 & 2.2 & 2.32 & 2.83 & 2.58 \\ 0.25 & 2.03 & 7.87 & 0 & 28.95 & 80.78 & 63.51 & 199.9 & 3.05 & 32.77 & 0 & 0 \\ 6.8 & 8.59 & 8.45 & 8.86 & 9.09 & 7.48 & 4.95 & 7.05 & 6.72 & 3.99 & 5.15 & 5.32 \end{bmatrix}$$

Get the following matrix according to the equation (2) to transform the value to a range.

$$X_{ij}^1 = \begin{bmatrix} 0.03 & 0.08 & 0.18 & 0.33 & 0.37 & 0.36 & 0.32 & 0.14 & 0.28 & 0.16 & 0.12 & 0.06 \\ 0.68 & 0.92 & 1 & 1 & 1 & 1 & 1 & 0.48 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0.93 & 0.74 & 0.79 & 0.5 & 0.54 & 0.18 & 0.59 & 0.57 & 0.57 & 0.6 \\ 0.02 & 0.02 & 0.03 & 0.08 & 0.04 & 0.04 & 0.03 & 0.01 & 0.03 & 0.01 & 0.03 & 0.03 \\ 0.02 & 0.01 & 0 & 0.04 & 0 & 0 & 0 & 0 & 0 & 0 & 0.04 & 0.04 \\ 0 & 0 & 0.08 & 0 & 0.36 & 0.99 & 0.66 & 1 & 0.01 & 0.32 & 0 & 0 \\ 0.06 & 0.1 & 0.09 & 0.12 & 0.08 & 0.06 & 0.03 & 0.03 & 0.05 & 0.02 & 0.06 & 0.07 \end{bmatrix}$$

Find the similarity matrix according to the equation (3).

$$R = \begin{bmatrix} 1 & 0.22 & 0.3 & 0.57 & 0.06 & 0.32 & 0.59 \\ 0.22 & 1 & 0.36 & 0.04 & 0.01 & 0.55 & 0.07 \\ 0.3 & 0.36 & 1 & 0.05 & 0.02 & 0.21 & 0.1 \\ 0.57 & 0.04 & 0.05 & 1 & 0.34 & 0.05 & 0.52 \\ 0.06 & 0.01 & 0.02 & 0.34 & 1 & 0 & 0.18 \\ 0.32 & 0.55 & 0.21 & 0.05 & 0 & 1 & 0.08 \\ 0.59 & 0.07 & 0.1 & 0.52 & 0.18 & 0.08 & 1 \end{bmatrix}$$

Seek the equivalent relation matrix according to the equation (4).

$$t(R) = \begin{bmatrix} 1 & 0.32 & 0.32 & 0.57 & 0.34 & 0.32 & 0.59 \\ 0.32 & 1 & 0.36 & 0.32 & 0.32 & 0.55 & 0.32 \\ 0.32 & 0.36 & 1 & 0.32 & 0.32 & 0.36 & 0.32 \\ 0.57 & 0.32 & 0.32 & 1 & 0.34 & 0.32 & 0.57 \\ 0.34 & 0.32 & 0.32 & 0.34 & 1 & 0.32 & 0.34 \\ 0.32 & 0.55 & 0.36 & 0.32 & 0.32 & 1 & 0.32 \\ 0.59 & 0.32 & 0.32 & 0.57 & 0.34 & 0.32 & 1 \end{bmatrix}$$

Draw a dynamic cluster diagram as figure 1.

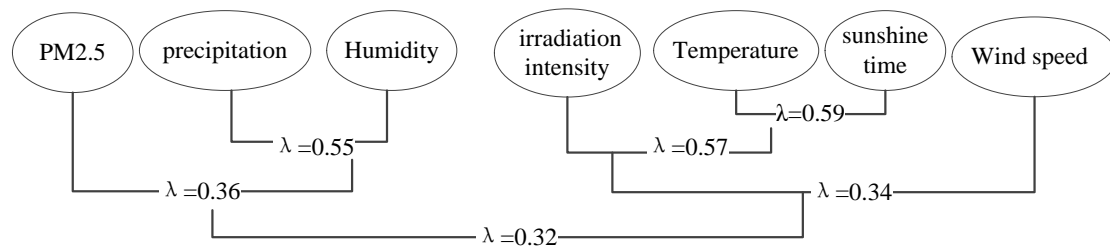


Figure 1. Dynamic Cluster Diagram

It is known from the dynamic cluster map that when >0.55 , the climate factors can be divided into four categories. The first category includes temperature, irradiation intensity and sunshine time. The second category obtains precipitation and humidity. The third category is PM2.5. The other one is the wind speed.

We choose the easily controlled factors in the test process to characterize the climatic factors as follows.

- Temperature
- Humidity
- PM2.5
- Wind speed

2.4. Conclusion

The environmental factors that affect the failure rate of the power module in the A area can be characterized by four indexes. They are temperature, humidity, PM2.5 and wind speed.

3. Analysis of Key Factors Based on Failure Rate

3.1. Grey T' s Correlation Degree Method

T' s correlation degree method is a kind of grey correlation degree methods, and revised Deng's correlation degree. Because of its symmetry, uniqueness, comparability and normalization, it is suitable for analysis of influence factors [8]. It can analysis the environmental impact factors of power failure, extract the key factors effectively, and provide theoretical basis for the design of power module, failure analysis and environmental test design.

3.1.1. Principle of Grey T' s Correlation Degree Method. The grey T' s correlation degree is calculated according to the degree of relative change of the time series curve of the factor. For discrete time series, the degree of approximation of the relative trend of the two curves is determined by the size of the increment after normalization the original variable during the two time series at every period of $\Delta t = \Delta t(k) - \Delta t(k-1)$ ($k=2,3,\dots,n$) [9]. The correlation degree of two time series is defined as the weighted average of the correlation coefficient between each period of $\Delta t(k)$ and the weight is $\Delta t(k)$.

3.1.2. Model of Grey T' s Correlation Degree Method

- Standardizing

The process of standardization makes up for the lack of comparability of grey correlation degree, making it possible to analysis the influencing factors [10].

$$D_0 = \frac{1}{n-1} \sum_{k=2}^n |X_0(t_k) - X_0(t_{k-1})| \quad (5)$$

$$D_j = \frac{1}{n-1} \sum_{k=2}^n |X_j(t_k) - X_j(t_{k-1})| \quad (6)$$

$$y_0 = \left\{ \frac{X_0(t_k)}{D_0}, k = 1, 2, 3, \dots, n \right\} \quad (7)$$

$$y_j = \left\{ \frac{X_j(t_k)}{D_j}, k = 1, 2, 3, \dots, n \right\} \quad (8)$$

- Seeking Incremental sequence

$$\Delta y_0 = \{\Delta y_0(t_k) = y_0(t_k) - y_0(t_{k-1}), k = 2, 3, \dots, n\} \quad (9)$$

$$\Delta y_j = \{\Delta y_j(t_k) = y_j(t_k) - y_j(t_{k-1}), k = 2, 3, \dots, n\} \quad (10)$$

- Calculation of correlation coefficient at each time period

$$\varepsilon(t_k) = \text{sgn}(\Delta y_0(t_k) \cdot \Delta y_j(t_k)) \cdot \frac{\min(|\Delta y_1(t_k)|, |\Delta y_j(t_k)|)}{\max(|\Delta y_1(t_k)|, |\Delta y_j(t_k)|)} \quad (11)$$

$$\varepsilon(t_k) = 0, \text{ when } \Delta y_0(t_k) \cdot \Delta y_j(t_k) = 0$$

In the upper equation, the positive and negative of correlation coefficient is determined by $\varepsilon(t_k)$. $\varepsilon(t_k) > 0$ when $\Delta y_0(t_k) \cdot \Delta y_j(t_k) > 0$ means that X_0 and X_1 change from t_{k-1} to t_k in the same direction, that is, positive correlation. $\varepsilon(t_k) < 0$ when $\Delta y_0(t_k) \cdot \Delta y_j(t_k) < 0$ means that X_0 and X_1 change from t_{k-1} to t_k in different directions, that is, negative correlation. $\varepsilon(t_k) = 0$ when $\Delta y_0(t_k) \cdot \Delta y_j(t_k) = 0$ means that X_0 and X_1 are irrelevant during $\Delta(t_k)$ [11].

3.1.3. Definition of Grey T' s correlation degree. The correlation degree between the original time series X_1 and X_2 is defined as $r(X_1, X_2)$.

$$r(X_1, X_2) = \frac{1}{b-a} \sum_{k=2}^n \Delta(t_k) \cdot \varepsilon(t_k) \quad (12)$$

3.2. Analysis of Key Factors

Through fuzzy clustering analysis, complex and changeable natural environment and climate factors are classified into different categories, and the categories of influencing factors are reduced. However, we can't clearly understand the impact of each type of natural environment and climate on the failure rate of the product. For the communication power module, temperature, humidity, wind and PM2.5 are the 4 dimensions that affect the failure rate. Based on the T correlation degree method, the correlation degree of the 4 dimensions to the power failure rate can be obtained.

Taking the power module of a certain communication equipment as an example, the quantitative calculation of the influence degree of natural environment factors based on the failure rate is carried out. The failure rate and environmental climate statistics are shown in Table 2.

Table 2. Failure Rate of a Certain Communication Power Module and Corresponding Environmental and Climatic Data

Time	Failure rate(10^{-6})	Temperature	Humidity	PM2.5	Wind speed
January	3.50	3.32	78.13	114.29	2.58
February	2.50	7.64	64.39	69.86	2.64
March	3.00	14.13	67.26	62.68	2.61
April	3.50	23.70	72.33	53.27	2.90
May	4.00	29.77	74.10	59.42	3.61
June	5.50	30.97	81.83	42.47	2.60
July	8.50	31.61	95.03	52.06	2.00
August	9.00	30.13	97.71	38.06	2.06
September	8.00	27.87	95.23	57.40	2.2
October	8.50	17.23	97.45	56.94	2.32
November	4.50	9.57	80.23	45.43	2.83
December	3.00	4.48	72.48	43.65	2.58

- Standardizing

According to equation (5) to (8), the result is shown in Table 3.

Table 3. Standardized Sequences

Time	Failure rate (10^{-6})	Temperature	Humidity	PM2.5	Wind speed
January	0.59	0.66	11.21	8.93	7.50
February	1.00	1.52	9.24	5.46	7.67
March	1.89	2.81	9.65	4.9	7.60
April	3.68	4.71	10.38	4.16	8.43
May	4.86	5.92	10.63	4.64	10.5
June	5.41	6.16	11.74	3.32	7.56
July	6.07	6.28	13.63	4.07	5.81
August	6.14	5.99	14.02	2.97	6.00
September	5.10	5.54	13.66	4.48	6.40
October	3.54	3.42	13.98	4.45	6.75
November	1.82	1.90	11.51	3.55	8.24
December	0.68	0.89	10.40	3.41	7.50

- Calculate increment sequence

The increment sequence is calculated according to equation (9) and (10), and the result is shown in Table 4.

Table 4. The Increment Sequence

Period	Failure rate (10^{-6})	Temperature	Humidity	PM2.5	Wind speed
1~2	0.41	0.86	-1.97	0.17	1.29
2~3	0.89	1.29	0.41	-0.08	1.63
3~4	1.79	1.9	0.73	0.83	1.56
4~5	1.18	1.21	0.25	2.07	0.74
5~6	0.54	0.24	1.11	-2.94	-0.46
6~7	0.66	0.13	1.89	-1.74	-1.22
7~8	0.07	-0.3	0.38	0.19	-0.43
8~9	-1.04	-0.45	-0.36	0.39	-0.53
9~10	-1.56	-2.12	0.32	0.36	-1.24
10~11	-1.72	-1.52	-2.47	1.48	-1.37
11~12	-1.13	-1.01	-1.11	-0.73	-0.57

- Calculate correlation degree

According to equation (11) and (12), the correlation coefficient and correlation degree of each period are calculated, as shown in Table 5.

Table 5. Correlation Coefficient and Correlation Degree in Each Period

Period	Failure rate(10^{-6})	Temperature	Humidity	PM2.5	Wind speed
1~2	1	0.48	-0.21	-0.12	0.42
2~3	1	0.69	0.46	-0.63	-0.09
3~4	1	0.94	0.41	-0.41	0.47
4~5	1	0.98	0.21	0.41	0.57
5~6	1	0.44	0.49	-0.41	-0.18
6~7	1	0.19	0.35	0.88	-0.38
7~8	1	-0.24	0.18	-0.06	0.38
8~9	1	0.43	0.34	-0.69	-0.38
9~10	1	0.74	-0.2	0.02	-0.23
10~11	1	0.88	0.7	0.52	-0.86
11~12	1	0.89	0.98	0.12	0.65
T' s correlation	1	0.58	0.34	-0.03	0.03

3.3. Results

The calculation results show the correlation degree between temperature, humidity, PM2.5 or wind speed and failure rate

- The grey T' s correlation degree of temperature and failure rate is 0.58.
- The grey T' s correlation degree of humidity and fault rate is 0.34.
- The grey T' s correlation degree of PM2.5 and fault rate is -0.03.
- The grey T' s correlation degree of wind speed and failure rate is 0.03.

Obviously, the correlation between temperature, humidity and failure rate is relatively large. However, the correlation between PM2.5 or wind speed and fault rate is very small.

4. Conclusions

In this paper, a method of reducing the dimension of environmental factors that affect product failure based on fuzzy clustering is proposed, and the grey T' s correlation degree analysis method is used to extract the key environmental factors that affect the product failure. Taking the power module of a certain type of communication equipment as an example, collecting and investigating related data and analyzing the main environmental factors that affect its failure are temperature and humidity.

According to 12 refs, 52% of the products caused by environment, 21% of which are temperature and 10% of humidity, are the main influencing factors [12]. This is in accordance with the conclusions of this paper, which illustrates the feasibility of the method described in this paper, thus laying a foundation for the analysis of the mechanism of product failure, the design of reliability test and the improvement of the reliability of the product.

References

- [1] S Z Song, G Y Wang, S Y Wang. Progress in Data Processing of Natural Environmental Corrosion of Materials in China *J. Journal of corrosion and protection of China*, 2003, **23** (1) pp 56-64.
- [2] Z J Jia, Q I Zhao, C F Dong, et al. The Principal Component Analysis of The Climate Factors of 13 Atmospheric Corrosion Sites in China *J. China Journal of corrosion and protection*, 2009, **29** (5) pp 388-393.
- [3] L W Zhang, X H Wang, X H Feng, et al. Fuzzy Clustering Analysis of Atmospheric Corrosion Environment and Environmental Factors *J. Chinese Journal of corrosion and protection*, 2003, **23** (2) pp 112-115.
- [4] J L Deng. The Basic Method of Grey System *M. Huazhong University of Science and Technology Press*, 1987.
- [5] M X Cao. Grey Relational Analysis Model and Its Application *D. Nanjing University of Aeronautics & Astronautics*, 2007.
- [6] J H Wu, X J Wang, Y Zhang. The Choice of Copula Function in Correlation Analysis *J. Statistical Research*, 2014, **31** (10) pp 99-107.
- [7] X Bai. Similarity Measurement and Application in Cluster Analysis *D. Beijing. JiaoTong University*, 2012.
- [8] Y G Sun, Y G Dang. Grey Grey T' s correlation degree Improved *J. System Engineering Theory and Practice*, 2008, **28** (4) pp 135-139.
- [9] M Tian, S F Liu, Z K Bu. Review of Grey Relational Algorithm Models *J. Statistics and Decision Making*, 2008, 2008 (1) pp 24-27.
- [10] W X Tang. Grey T' s correlation degree and Its Calculation Method *J. Mathematical Statistics and Management*, 1995 (1) pp 34-37.
- [11] Z G Mei. Grey Absolute Correlation Degree and Its Calculation Method *J. System Engineering*, 1992, **10** (5).
- [12] Y T Gao, Y Xu, J Chang. Multi Environmental Test of Sight and Theory of Thermal Performance of Infrared Optical Materials *M. Tsinghua University press*, 2017.