# *Composition analysis of ancient glass products based on logistic regression and principal component analysis*

**Tianyu Zheng[1,#], Ziyi Gong[1,#], Yutong Chen[2,#], Qianyun Ma[3]**

*[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China*
*[2]School of Science, Wuhan University of Science and Technology, Wuhan, China*
*[3]School of Foreign Languages, Wuhan University of Science and Technology, Wuhan, China*
*[#]These authors contributed equally to this work*

*Abstract:* In order to analyze and study the composition of ancient glass products, this paper preprocessed the data and carried out statistical analysis of charts to qualitatively analyze the relationship between weathering and color, type and pattern of cultural relics, and established a Logistic regression model for quantitative analysis. There are four types of cultural relics according to the type of cultural relics and whether they are weathered. Through the evaluation model of principal component analysis, the statistical law of weathering and chemical composition content of excavated glass cultural relics is calculated according to the corresponding comprehensive score range of each kind. A multiple linear regression model was established to predict the pre-weathering component content. The correlation and difference between the chemical components of different kinds of glass relics were analyzed. The correlation coefficients of high potassium glass and lead barium glass were analyzed respectively, and the correlation coefficient heat maps were drawn. Determine the relationship between the chemical components and the differences between different chemical components.

## 1. Introduction

The Silk Road was a channel for cultural exchanges between China and the West in ancient times, and glass is valuable evidence of early trade. From the 10th to the 12th century AD, glass cultural relics were gathered in China at their peak, and the unearthed glass cultural relics in this period were widely distributed and culturally distinctive[1]. China's ancient glass products and foreign glass products similar in appearance, but the chemical composition is not the same. The main chemical component of glass is silicon dioxide ($SiO_2$). However, the melting point of pure quartz sand is high, and in order to reduce the melting temperature, flux and stabilizer need to be added in refining. The main chemical composition of glass is also different if the flux added to the glass is different[2]. The limestone added during refining can act as a stabilizer to prevent weathering of glass, but the chemical composition ratio of ancient glass will change after weathering[3]. Ancient glass is susceptible to weathering under the influence of burial environment.

In the process of weathering, the internal elements have a lot of exchange with the environmental elements, resulting in a change in their composition ratio, which affects the correct judgment of their categories. Different degrees of weathering, the content of elements inside the glass is not the same, and the color of the glass is not the same.

The purpose of this paper is to explore the relationship between weathering, color, type and decoration of glass cultural relics by means of data preprocessing and chart statistical analysis, and at the same time to classify four types of cultural relics by logistic regression model. Through the evaluation model of principal component analysis, the comprehensive score range of various cultural relics was calculated, so as to deeply understand the weathering condition and chemical composition content law. The multivariate linear regression model is also established, which refers to the maximum principal component quantity by variance explanation and cumulative projection importance. Then the composition of principal components is obtained through component matrix table. Then the importance of variables is obtained through the factor load coefficient table. Finally, the normalized formula of weathering degree was obtained by partial least squares regression[4], and the pre-weathering component content was predicted by monitoring point data. On this basis, correlation coefficient analysis[5] and thermal map mapping were used to analyze the correlation and difference between the chemical components of different types of glass cultural relics, so as to determine the correlation and difference level between the components. Through these professional research and analysis, we hope to better understand the protection and restoration of cultural relics, and provide a reliable scientific basis for the identification, restoration and preservation of cultural relics under the modern scientific and technological environment.

## 2. Model building and solving

### 2.1. Based on Logistic regression prediction model

Observing the chemical composition ratio data of classified glass relics, in theory, the sum of the proportion of each component should be 100%, but due to detection methods and other reasons, the sum of the proportion of its composition and non-100% situation. Data with the sum of all the chemical components of ancient glass between 85% and 105% are used as qualified data[6]. Data with the sum of chemical components outside this range are regarded as abnormal data and eliminated. A null value in the data indicates that the component is not detected by the detection means and is treated as 0.
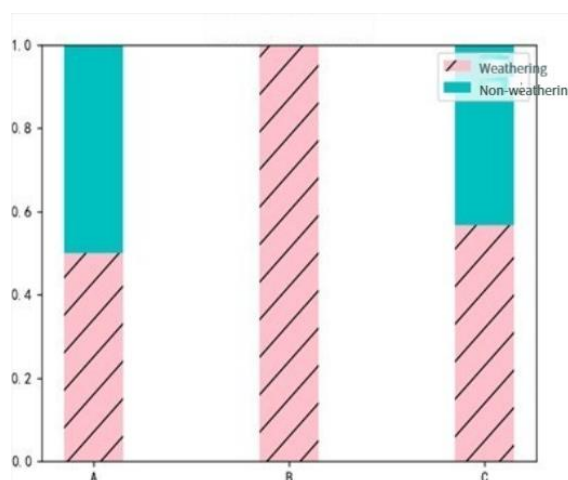


Figure 1: The proportional relationship between glass ornamentation and weathering.
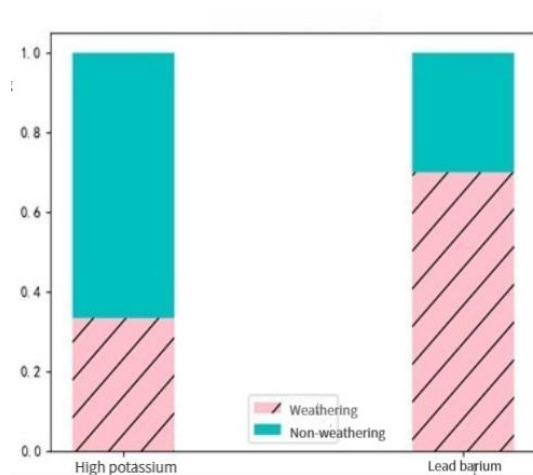
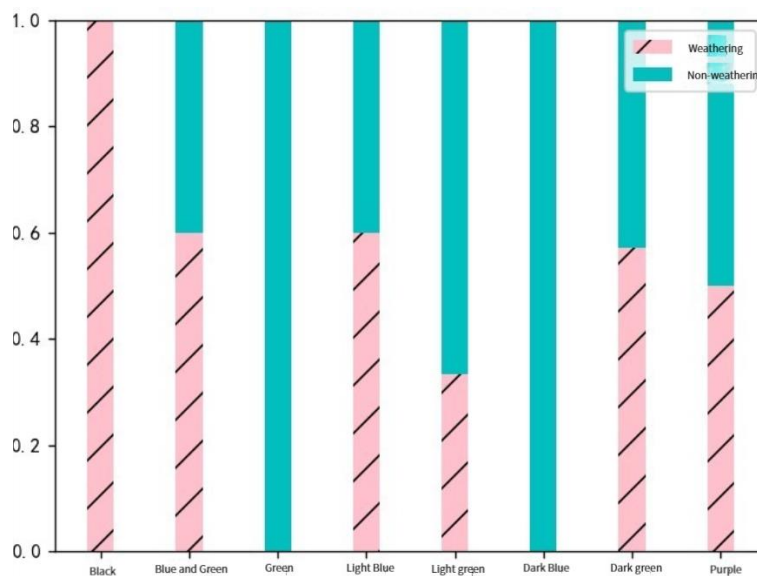Figure 2: Relationship between glass type and weathering ratio.



Figure 3: The relationship between glass color and weathering ratio.

The relationship between glass decoration and weathering can be seen from Figure 1. The glass relics decorated with B decoration are more easily weathered than those decorated with A decoration and C decoration, and the weathering conditions of glass relics decorated with A decoration and C decoration are similar. Figure 2 shows the relationship between glass types and weathering. Lead-barium glass is more susceptible to weathering than high-potassium glass, and the proportion of lead-barium glass to weathering is higher than that of potassium glass. Figure 3 shows the relationship between glass color and weathering. Black cultural relics are basically weathered, while dark blue and green cultural relics are basically unweathered. The proportion of green unweathered was higher than that of dark green and light green, and the proportion of light green unweathered was higher than that of dark green. Dark blue has a higher proportion of unweathered weather than purple, light blue and turquoise.

Logistic regression model was established to find out the internal rule of the quantitative relationship between glass weathering and glass type, decoration and color. Firstly, the relationship between weathering and a single variable is considered, that is, the Logistic equation of glass type, decoration and color. Then consider the relationship between weathering and two variables, that is, discuss whether there is an interaction between a single variable on "weathering". Established with

or without weathering and glass types and ornamentation; There is no weathering and glass type and color; There are three Logistic equations of weathering pattern and color; Finally, Logistic equations of weathering and glass type, ornamentation and color are established. A total of seven Logistic mathematical models were obtained, as shown in Table 1.

Table 1: This caption has one line so it is centered.

| ID | Variable | Logistic model |
|----|----------|----------------|
| 1 | color | $\ln(\frac{p}{1-p}) = \alpha + \sum_{i=1}^{7} \beta_i x_i$ |
| 2 | ornamentation | $\ln(\frac{p}{1-p}) = \alpha + \sum_{j=1}^{2} \varepsilon_j y_j$ |
| 3 | species | $\ln(\frac{p}{1-p}) = \alpha + \sum_{k=1}^{1} \gamma_k z_k$ |
| 4 | color & ornamentation | $\ln(\frac{p}{1-p}) = \alpha + \sum_{i=1}^{7} \beta_i x_i + \sum_{j=1}^{2} \varepsilon_j y_j$ |
| 5 | color & species | $\ln(\frac{p}{1-p}) = \alpha + \sum_{i=1}^{7} \beta_i x_i + \sum_{k=1}^{1} \gamma_k z_k$ |
| 6 | Ornamentation& species | $\ln(\frac{p}{1-p}) = \alpha + \sum_{j=1}^{2} \varepsilon_j y_j + \sum_{k=1}^{1} \gamma_k z_k$ |
| 7 | color & ornamentation& species | $\ln(\frac{p}{1-p}) = \alpha + \sum_{i=1}^{7} \beta_i x_i + \sum_{j=1}^{2} \varepsilon_j y_j + \sum_{k=1}^{1} \gamma_k z_k$ |

As shown in Figure 4, it can be found that the type of glass cultural relics has the greatest influence on predicting the weathering probability of glass cultural relics. The Logistic equation based on glass type, color and pattern predicted the weathering results with the highest accuracy.
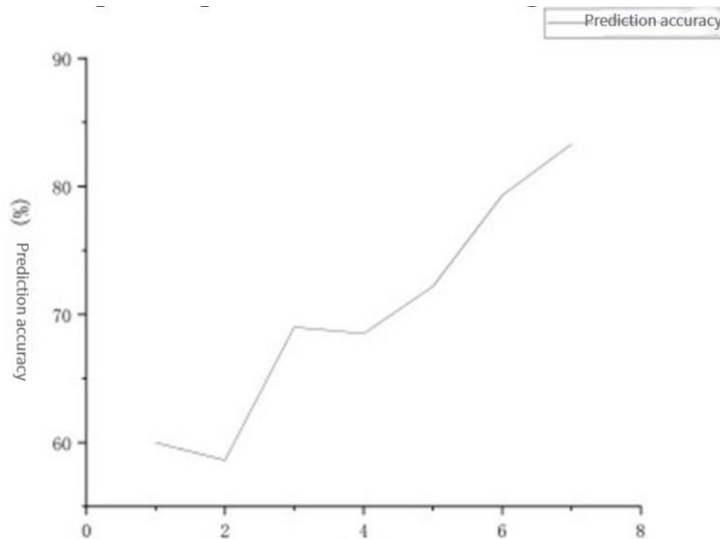


Figure 4: Weathering prediction accuracy of 7 Logistic equations.

## 2.2. Evaluation model based on principal component analysis

According to the types of cultural relics and whether they are weathered, they are divided into

four types: weathered high-potassium glass, non-weathered high-potassium glass, weathered lead-barium glass, and unweathered lead-barium glass; Assume that there are $m$ variables for principal component analysis and a total of $n$ evaluation objects, $x_{ij}$ is the jth index of the ith evaluation object, and the value of each index $x_{ij}$ is converted into a standardized index $\tilde{x}_{ij}, i = 1, 2, \cdots, m$ is the sample mean value of the jth indicator, and $s_j$ is the standard deviation of the jth indicator, as shown in Equation (1).

$$\tilde{x}_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j} \tag{1}$$

Calculate the correlation coefficient matrix R, where $R_{ij}$ is the correlation coefficient of the jth index of the i index, as shown in equation (2)(3).

$$r_{ij} = \frac{\sum_{k=1}^{n} \tilde{x}_{ki} \tilde{x}_{kj}}{n-1} (i, j = 1, 2, \cdots, m) \tag{2}$$

$$R = (r_{ij})_{m \times n} \tag{3}$$

Calculating the eigenvalues of the correlation coefficient matrix R $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$ and eigenvector $u_1, u_2, \cdots, u_j$, including $u_j = (u_{1j}, u_{2j}, \cdots, u_{nj})^T$, has a characteristic vector of m a new indicator variables. As shown in equation (4), where $y_1$ is the first principal component, $y_2$ is the second principal component, and $y_m$ is the m principal component.

$$\begin{cases} y_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \cdots + u_{n1}\tilde{x}_n \\ y_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \cdots + u_{n2}\tilde{x}_n \\ \qquad \vdots \\ y_m = u_{1m}\tilde{x}_1 + u_{2m}\tilde{x}_2 + \cdots + u_{nm}\tilde{x}_n \end{cases} \tag{4}$$

n(n≤m) principal components were selected to calculate the comprehensive evaluation value. The information contribution rate and cumulative contribution rate of the eigenvalue $\lambda_j (j = 1, 2, \cdots, m)$ are calculated, as shown in equation (5). The information contribution rate of principal component $y_i$ is calculated as shown in equation (6).

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{m} \lambda_k} (j = 1, 2, \cdots, m) \tag{5}$$

$$a_p = \sum_{k=1}^{p} \lambda_k / \sum_{k=1}^{m} \lambda_k \tag{6}$$

By analyzing the variance interpretation, the contribution rate $\alpha_p$ is calculated, and the number of principal components is determined as n according to the principle that the contribution rate reaches 85%. By calculating and analyzing the principal component load coefficient matrix G, the importance of hidden variables in each principal component can be obtained. The composition matrix H of the four types of glass was calculated and analyzed respectively, as shown in equation

(7). The coefficients in the component matrix are linear combination coefficients, and the principal component scoring formula and the total scoring formula are obtained.

$$H = \frac{G}{\sqrt{\lambda}} \tag{7}$$

$$a_j = b_j / \alpha_p \tag{8}$$

$$F_i = \sum_{j=1}^{m} b_j Q \tag{9}$$

$$F = \sum_{i=1}^{m} a_i F_i \tag{10}$$

The principal component weight is α_j, as shown in equation (8); Where $F_i$ is the score of the i th principal component, $b_j$ is the linear combination coefficient, and Q is the content of each differentiation component of a cultural relic. Where F is the total score of a certain cultural relic, and $a_i$ is the weight of the main component.

The characteristic range of weathered high-potassium glass, non-weathered high-potassium glass, weathered lead-barium glass, and unweathered lead-barium glass is calculated, as shown in Table 2 below.

Table 2: Characteristic range

| Type | Calculation formula | Feature range |
|------|---------------------|---------------|
| Unweathered high potassium glass | $F_a = (0.288/0.883) \times F_1 + (0.245/0.883) \times F_2 + (0.146/0.883) \times F_3$ $+ (0.122/0.883) \times F_4 + (0.081/0.883) \times F_5$ | [4.063,5.523] |
| Weathered high potassium glass | $F_b = (0.514/0.89) \times F_1 + (0.376/0.89) \times F_2$ | [-14.216,-13.954] |
| Unweathered lead barium glass | $F_c = (0.272/0.897) \times F_1 + (0.221/0.897) \times F_2 + (0.159/0.897) \times F_3$ $+ (0.108/0.897) \times F_4 + (0.072/0.897) \times F_5 + (0.065/0.897) \times F_6$ | [-5.376,0.611] |
| Weathered lead barium glass | $F_a = (0.281/0.876) \times F_1 + (0.222/0.876) \times F_2 + (0.124/0.876) \times F_3$ $+ (0.069/0.876) \times F_4 + (0.069/0.876) \times F_5 + (0.063/0.876) \times F_6$ $+ (0.046/0.876) \times F_7$ | [-1.974,3.091] |

## 2.3. Multiple linear regression prediction model

$$Y_f = K \cdot \beta + \varepsilon \tag{11}$$

$$K' = Y_f \cdot A + N_i \cdot B + \varepsilon' \tag{12}$$

$$K = (k_{ij})_{69 \times 14} \tag{13}$$

$$K' = (k'_{ij})_{67 \times 14} \tag{14}$$

As shown in equation (11), $Y_f = (y_1, y_2, \cdots, y_i)^T$ says the weathering degree of the vector of all monitoring, coefficient vector $\beta = (\beta_1, \beta_2, \cdots, \beta_i)^T$. As shown in equation (12), $A = (a_1, a_2, \cdots, a_{14})$, $B = (b_1, b_2, \cdots, b_{14})$ are the coefficient matrix, sample number N is introduced to reduce the influence of content differences between different samples, and $\varepsilon'$ is the constant term. As shown in equation (13), $k_{ij}$ represents the content of the jth chemical component at the i monitoring point. As shown in equation (14), $k_{ij}'$ represents the content of the jth chemical component at the i monitoring point after prediction.

According to the monitoring points, the contents of each chemical component before weathering of all cultural relics were obtained, as shown in Figure 1, and the distribution of each component content of the three cultural relics glass was predicted. It can be seen that the contents of PbO and BaO will increase significantly with the higher the degree of cultural relics weathering; The contents of K2O, CaO, CuO, P2O5, SrO and SO2 will increase with the higher degree of cultural relics weathering. The contents of Na2O, MgO, Al2O3 and Fe2O3 will decrease with the higher degree of cultural relics weathering. With the higher the degree of weathering of cultural relics, the content of SiO2 will be significantly reduced.
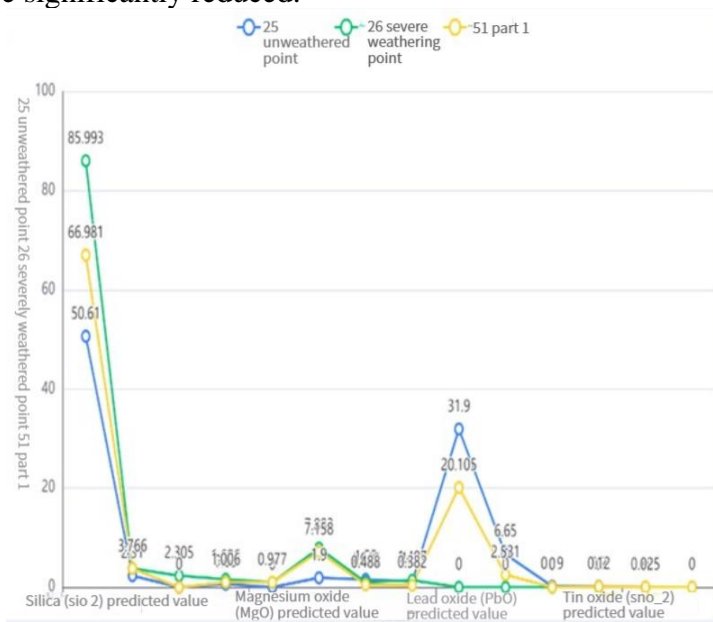


Figure 5: Predicted composition content of 3 kinds of cultural glass

## 2.4. Multiple linear regression prediction model

Pearson correlation analysis method was adopted, as shown in equation (15). By calculating the chemical components x and y to measure the degree of linear correlation between them, the Pearson correlation coefficient r is obtained.

$$r(x, y) = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{15}$$

First, the data in the integration table is divided into two categories according to the type, and the correlation coefficient matrix between each compound is obtained. The two-dimensional thermal map of the correlation coefficient is drawn with the coefficient matrix. The darker the color, the

closer the value of the correlation coefficient is to 1, indicating the stronger the positive correlation, and the lighter the color, the closer the correlation coefficient is to -1. The stronger the negative correlation is. The color is centered, and the closer the data is to 0, the weaker the correlation. Data with the absolute value of correlation coefficient greater than 0.9 are used as the judgment criteria for correlation. By analyzing this figure, the correlation between the chemical components of different types of glass cultural relics samples was obtained, and the differences between the correlation coefficients of different types of chemical components were compared.

By comparing Figure 6 and Figure 7, it can be found that the color shades in Figure 6 are clearly divided, while those in Figure 7 are more evenly divided. It shows that there are more connections between the chemical components of lead barium glass and less connections between the chemical components of high potassium glass.

The differences of chemical composition correlation between different types of glass are as follows: 1. There is a strong positive correlation between alumina and silica in lead-barium glass, but a strong negative correlation between alumina and silica in high-potassium glass; 2. There is a strong negative correlation between strontium oxide and phosphorus pentoxide in lead-barium glass, and a strong positive correlation between strontium oxide and phosphorus pentoxide in high-potassium glass. 3. In general, the correlation between the chemical components of lead barium glass is mostly stronger than that between the chemical components of high potassium glass. However, due to the small number of samples of high-potassium glass, it may weaken the correlation between the chemical components of high-potassium glass, so this difference is not particularly obvious compared with the first two.
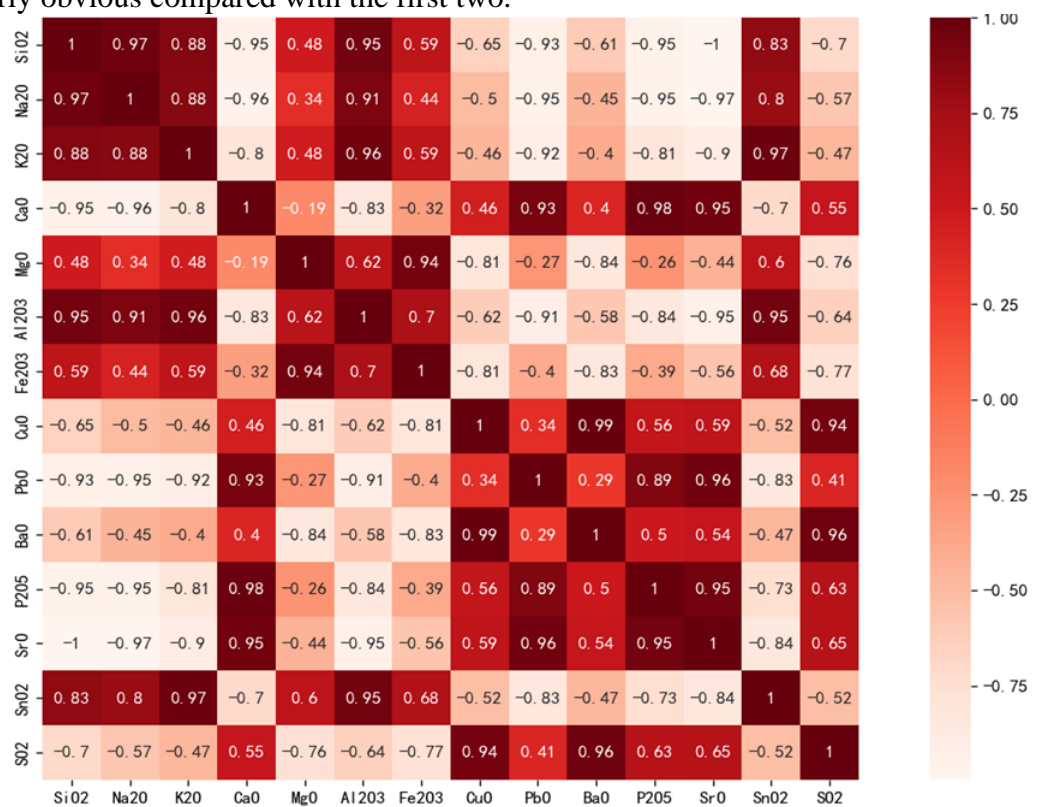


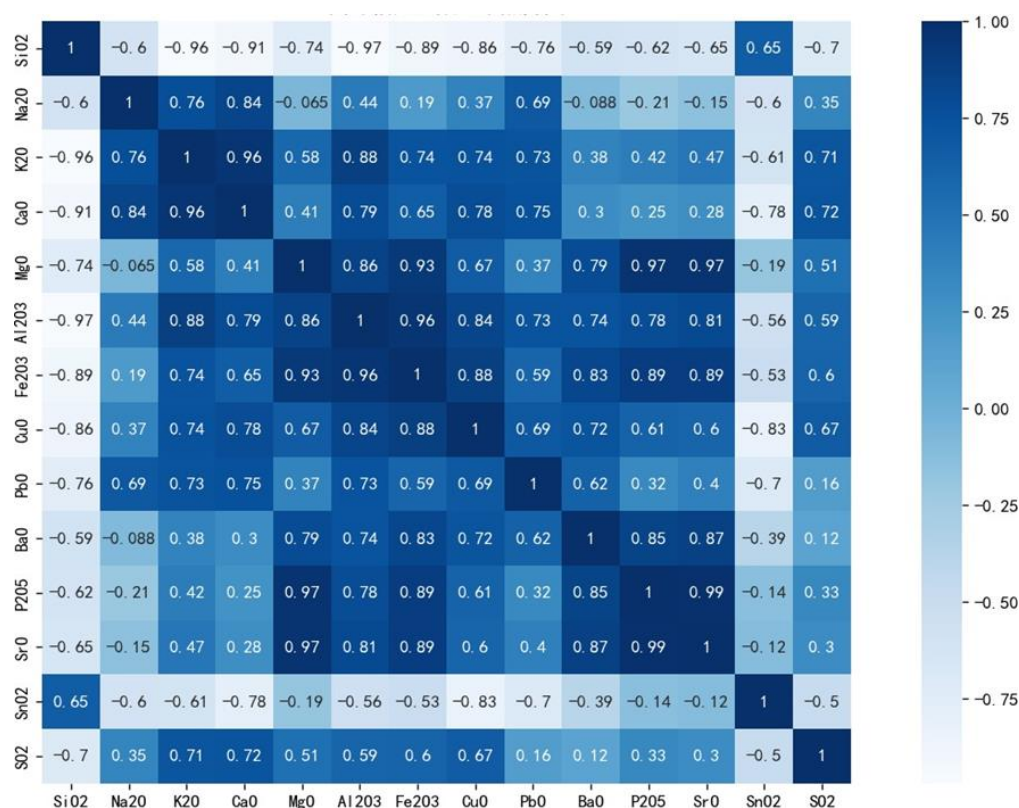Figure 6: Lead barium correlation coefficient two-dimensional heat map

Figure 7: High potassium correlation coefficient two-dimensional heat map

## 3. Conclusions

In this paper, principal component analysis can eliminate the correlation between each chemical component and reduce the workload of selecting indicators. The correlation coefficient method can be used to obtain the correlation effect of each chemical component more directly. This paper uses a variety of analysis methods, such as data preprocessing, chart statistical analysis, logistic regression model, principal component analysis and multiple linear regression model, to deeply study multiple factors of glass relics from different angles. The logistic regression model was used to classify four types of cultural relics. The principal component analysis method was used to evaluate the model, and the comprehensive score range of various cultural relics was determined. The correlation and differences among the chemical components of different types of glass cultural relics were deeply analyzed through correlation coefficient analysis and thermal map rendering, providing reference value for cultural relics identification, restoration and preservation. The model can be applied to various types of cultural relics, and the result is universal and extensive, which is a research of practical value. However, the data sources in this paper may be limited, and the sample size is limited, which may affect some results. Although this paper adopts a variety of methods to establish the model, the correlation coefficient method cannot judge the strong correlation relationship with nonlinear, and the model itself is uncertain, which cannot accurately predict the color, type, chemical composition and other related factors of cultural relics, so it needs to be combined with the field observation and judgment of professionals.

The model presented in this paper can be applied to the related fields of the protection, restoration and reconstruction of glass cultural relics, expand the number of samples, and more comprehensively understand the degree of weathering and chemical composition of cultural relics. The model can be applied to the actual identification, restoration and preservation of cultural relics,

and provide scientific basis for the protection and restoration of cultural relics. At the same time, the research method and analysis results of this paper can also be applied to other types of cultural relics research, expand the application range, and improve the academic value and practicability of the research results.

## References

[1] Li Yingli. Research on Glassware unearthed in China from the 10th to the 12th century [D]. Liaoning: Liaoning Normal University, 2021.

[2] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1. 0. 11) [Online Application Software]. Retrieved from https://www.spsspro.com. 1-2.

[3] Gu Liangliang, Zhou Jing. Some Problems on the Research of Ancient Chinese Glass Art [J]. Art and Design, 2021, 2(8):111-114.

[4] Liang Beichen, Dai Jingmin. Application of Partial least squares Method in System Fault diagnosis [J]. Journal of Harbin Institute of Technology, 20, 52(3):156-164. DOI:10. 11918/201805149. 1-2.

[5] He Xiaoqun. Multivariate statistical analysis. Beijing: China Renmin University Press, 2012. 1-2.

[6] Gan Fuxi, Zhao Hongxia, Li Qinghui, et al. Science and technology analysis and research of the Warring States Period glass unearthed in Hubei Province [J]. Jianghan Archaeology, 2010(2):108-116, 151, 152.