Analytical Chemistry: A Journal (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/analc.2022.010109
ISSN 2616-1958 Vol. 1 Num. 1

Establishment and application of inferential equations for glass types

Wenbo Wei, Haiyu Wang, Runze Yan[*]

College of Mathematical Science and Engineering, Hebei University of Engineering, Handan, Hebei, 056038, China
[*]Corresponding author: runzeyan2002@163.com

Keywords: Gray correlation, Principal component analysis, Inference equation

Abstract: In order to analyze the classification rules of high-potassium glass and lead-barium glass, this paper performs gray correlation analysis on the two types of glass, selects the appropriate chemical composition for each category to divide them into subcategories, gives specific division methods and division results according to the results of principal component analysis, and derives the inferred equation for each category of glass. The inferred equation is used to analyze the chemical composition of the unknown category of glass artifacts, calculate the central distance, and determine which glass category it is by comparing the size of the central distance Identify the type to which it belongs.

1. Introduction

Glass was an important item in the early trade between China and the West, and because of the scarcity of ancient glass products [1-3], it has a very high collector's value and artistic value. The chemical composition of ancient glass differs from that of Western glass due to the geographical and material conditions in China [4-6]. Therefore, the analysis of the chemical composition of ancient glass objects is an important task, and archaeologists can infer and identify the region, age, production method [7], and degree of weathering based on the chemical composition of glass. The analysis of the composition of ancient glass objects is of great significance for the study of ancient history and the conservation of cultural relics [8-10].

2. Establishment of inferential equations for glass types

2.1 Modeling by principal component analysis

Assuming a total of n samples and p indicators in the sample, a sample matrix x of size n × p can be formed.

$$x=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}=(x_1,x_2,\cdots,x_p) \quad (1)$$

Standardize the sample and calculate the correlation coefficient matrix Calculate mean values by column:

$$\bar{x}_j = \frac{1}{n}\sum xij \qquad (2)$$

Standard deviation:

$$S_j = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}{n-1}} \qquad (3)$$

The standardized data is:

$$X_{ij} = \frac{x_{ij}-\bar{x}_j}{S_j} \qquad (4)$$

After standardization the sample matrix changes to:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = (X_1, X_2, \cdots, X_p) \qquad (5)$$

Calculate the correlation coefficients for each row and column:

$$r_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ki}-\bar{X}_i)(X_{ki}-\bar{X}_j) = \frac{1}{n-1}\sum_{k=1}^{n}X_{ki}X_{kj} \qquad (6)$$

After processing the above analysis, the correlation coefficient matrix can be obtained as:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \qquad (7)$$

Calculate the eigenvalues and eigenvectors:

Eigenvalues of the correlation coefficient matrix R $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$.

Corresponding eigenvectors:$u_1, u_2, \cdots, u_m$, Among them $u_j = (u_{1j}, u_{2j}, \cdots, u_{rj})^T$, m new indicator variables composed of feature vectors

$$\begin{cases} y_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \cdots + u_{n1}\tilde{x}_n \\ y_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \cdots + u_{n2}\tilde{x}_n \\ \cdots \\ y_m = u_{1m}\tilde{x}_1 + u_{2m}\tilde{x}_2 + \cdots + u_{nm}\tilde{x}_n \end{cases} \qquad (8)$$

wherey_1 is the first principal component, y_2 is the second principal component, ..., $y_m$is the mth principal component

Calculating Eigenvalues $\lambda_j(j = 1,2,\cdots,m)$ the information contribution rate and cumulative contribution rate of

$b_j = \frac{\lambda_j}{\sum_{k=1}^{m}\lambda_k}(j = 1,2,\cdots,m)$ is the information contribution of the principal component y_j;

$\alpha_p = \frac{\sum_{k=1}^{p}\lambda_k}{\sum_{k=1}^{m}\lambda_k}$ is the cumulative contribution of the principal component y_j;

Calculate the composite score and regression coefficient of each principal component

$$Z = \sum_{j=1}^{p} b_j y_j \qquad (9)$$

where b_j is the information contribution of the jth principal component.

The regression coefficients are calculated based on the scores of the principal components:

$$xs = \frac{z}{x} \tag{10}$$

Z is the composite score and x is the normalized sample matrix.

Calculate the coefficients of the regression equation of the original variables

The coefficients of the regression equation of the original variables can be calculated from the mean and standard deviation of the samples, and the regression equation of the speculative glass model can be established.

2.2 Analysis of the classification law of high potassium glass and lead-barium glass

The combined table was split into a high potassium glass table and a lead-barium glass table to analyze the gray correlations of high potassium glass and lead-barium glass separately. The chemical components with the top four gray correlations (the reasons for selecting the four components are given below in the principal component analysis) were selected for the analysis. The results are shown in the table 1 and 2.

Table 1: Grey correlation analysis of high potassium glass

	Correlation result	
Evaluation items	Relevancy	Ranking
SiO2	0.981	1
Al2O3	0.945	2
CuO	0.939	3
P2O5	0.921	4

Table 2: Grey correlation analysis of lead barium glass

	Correlation result	
Evaluation items	Relevancy	Ranking
PbO	0.967	1
SiO2	0.964	2
BaO	0.955	3
Al2O3	0.954	4

According to the above two figures, it can be seen that the chemical composition of silicon dioxide (SiO2) is higher when the glass type is high potassium glass, and the content of lead oxide (PbO) is higher when the glass type is lead-barium glass. High potassium glass is mainly related to the content of silicon dioxide (SiO2), aluminum oxide (Al2O3), copper oxide (CuO), and phosphorus pentoxide (P2O5). Lead-barium glass is mainly related to the content of lead oxide (PbO), silicon dioxide (SiO2), barium oxide (BaO), aluminum oxide (Al2O3).

2.3 Perform subcategorization

According to the above gray correlation analysis of high potassium glass and lead-barium glass respectively, it can be seen. High potassium glass is mainly related to silica, alumina, copper oxide and phosphorus pentoxide. The lead-barium glass is mainly related to lead oxide, silica, barium oxide, and alumina.

The specific classification of each glass type is given by the following principal component analysis of the data in the combined table.

Firstly, the principal component analysis was carried out by the principal component of 1 to obtain the following fragmentation diagram. By observing the fragmentation diagram, we can see that the image tends to be flat when the number of principal components k=4, so the number of principal components can be determined as 4. The results are shown in the figure 1

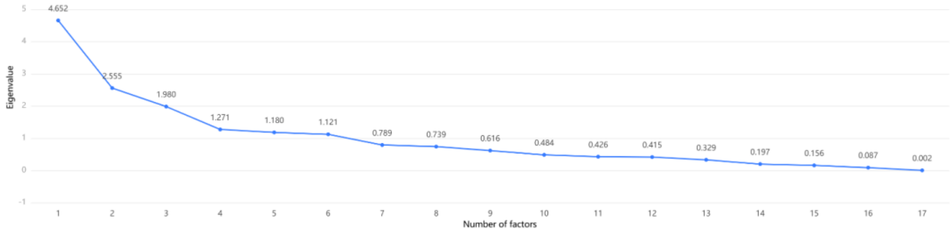

Figure 1: Factor analysis gravel plot

The four main components were identified as lead oxide (PbO), silicon dioxide (SiO2), barium oxide (BaO), and aluminum oxide (Al2O3) based on the magnitude of the correlation through the gray correlation analysis of the various chemical components and glass types in the combined table 3.

Table 3: Correlation magnitudes of the four principal components

	Correlation result	
Evaluation items	Relevancy	Ranking
PbO	0.966	1
SiO2	0.963	2
BaO	0.96	3
Al2O3	0.96	4

Solve the principal component regression equation for the type of glass:

(1) Let lead oxide be x1, silicon dioxide be x2, barium oxide be x3, alumina be x4, and glass type be y1.

(2) Calculate the coefficient of ordinary least squares for y1 as
[0.029499 0.013397 0.036319 0.053058]

(3) and derive the correlation coefficient matrix A as

$$\begin{bmatrix} 1.0000 & -0.8418 & 0.3282 & -0.3914 \\ -0.8418 & 1.0000 & -0.6169 & 0.2321 \\ 0.3282 & -0.6169 & 1.0000 & -0.3378 \\ -0.3914 & 0.2321 & -0.3378 & 1.0000 \end{bmatrix}$$

(4) Calculate the eigenvalue of A as:

$$\lambda = \begin{bmatrix} 2.4246 \\ 0.8280 \\ 0.6843 \\ 0.0631 \end{bmatrix}$$

The eigenvector is obtained as:

$$\text{vec} = \begin{bmatrix} -0.5539 & 0.2065 & -0.5381 & 0.6009 \\ 0.5882 & -0.3950 & 0.0303 & 0.7051 \\ -0.4674 & -0.0558 & 0.8209 & 0.3234 \\ 0.3588 & 0.8934 & 0.1890 & 0.1930 \end{bmatrix}$$

(5) The proportion of each principal component (after ranking) was calculated from the eigenvalues as:

$$\text{rate} = \begin{bmatrix} 60.6139 \\ 20.7012 \\ 17.1071 \\ 1.5778 \end{bmatrix}$$

By the ratio it is found that the effects are the same so it can be assumed that the principal component fraction is 4.

(6) Calculate the regression coefficients of the principal component variables:

$$xs = \begin{bmatrix} 0.4743 \\ 0.2654 \\ -0.0401 \\ 1.5806 \end{bmatrix}$$

Finally, we get the regression coefficient of the original variable.

(7) The final principal component regression.

(8) Calculate the residuals and judge the reasonableness of the equation

The residual standard deviation of the n-1 parameters is: 0.2264.

The remaining standard deviation of the sum parameters is: 0.2228.

The principal component regression equation can be judged as reasonable from the two residuals.

From the above analysis, it can be seen that the classification of glass type by principal component analysis, the classification results in the inferred equation of glass type by four main components.

2.4 Analysis of the reasonableness and sensitivity of the classification results

The regression equation obtained by the principal component analysis is inferred from the main chemical components in the combined table for the glass types. The rate of change is 0.18, which can be regarded as reasonable because of the small rate of change. The results are shown in the table 4.

Table 4: Variation in glass type

	High Potassium	Lead Barium
Original number	18	49
Processed data	30	37

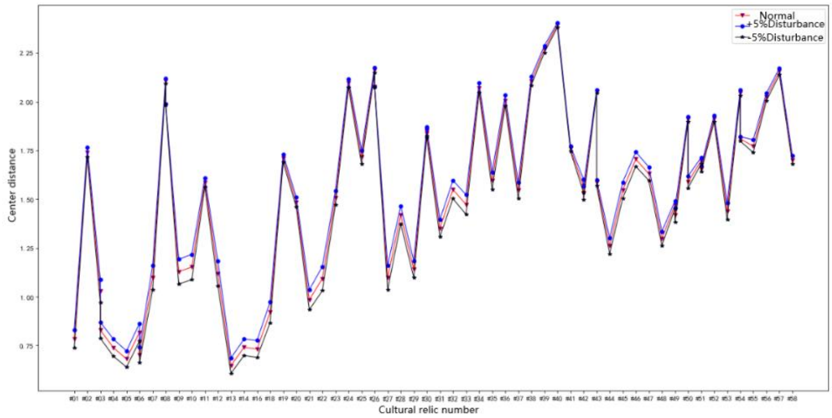

Figure 2: Effect of ±5% data perturbation on silica on center distance

Sensitivity analysis can analyze the robustness of the model by perturbing the data for any of the principal components. The perturbed principal component content is then used to infer the glass type based on the regression equation of the principal component analysis (the data are obtained as the center distance, if the distance to 1 is close, it is high potassium glass, if not, it is lead-barium glass, if the distance to 2 is closer, it is lead-barium glass, otherwise, it is high potassium glass). In this paper, the silica content is perturbed by ±5%, and the type of glass does not change when the data is perturbed, as can be seen from the line graph of center-to-center distance below. The results are shown in the figure 2.

3. Identification of unknown glass

3.1 Analysis of chemical composition of glass artifacts of unknown category

Using the regression equation for the glass type established by the principal component analysis in Problem 2, the center distance is calculated by combining the four principal components of lead oxide (PbO), silicon dioxide (SiO_2), barium oxide (BaO), and aluminum oxide (Al_2O_3) in Form 3 and determining whether the center distance is close to 1 or close to 2. If it is close to 1, it is a high potassium glass, and if it is close to 2, it is a lead-barium glass. Therefore, the glass artifacts of unknown category in Form 3 belong to the following glass types. The results are shown in the table 5.

Table 5: The glass artifacts of unknown category belong to the following glass types

Artifact Number	Center Distance	Prediction Type
A1	1.2906	High Potassium
A2	1.4972	High potassium
A3	1.7774	Lead Barium
A4	1.7244	Lead barium
A5	1.8333	Lead barium
A6	1.1848	High potassium
A7	1.3413	High potassium
A8	1.6918	Lead barium

3.2 Analysis of the sensitivity of the classification results

Based on the robustness, the data is perturbed for any of the main components and the center distance is calculated to determine whether the glass type changes. In this paper, the glass type does not change when the data is perturbed by ±5% for silica content, as can be seen from the following line graph.The results are shown in the figure 3.

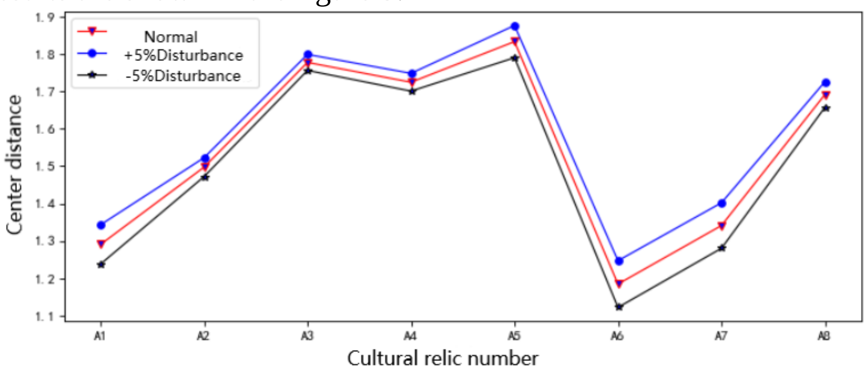

Figure 3: Perturbation of ±5% of the data for silica on

4. Conclusions

In this paper, gray correlation analysis was performed for high potassium or lead-barium glasses, and the classification pattern was analyzed according to the correlation magnitude of each chemical composition. The inference equation of each glass category was obtained by using Matlab to determine the number of principal components in the gravel diagram and to find the major components according to the correlation magnitude. The results were checked for reasonableness and sensitivity analysis for data perturbation based on robustness. The correlation magnitudes were compared to show that high potassium glasses were used when the SiO2 content was high, and lead-barium glasses were used when the PbO content was high. The number of principal components was 4. The four principal components with higher correlation were selected to classify them into subclasses. The inferred equation of the glass type was derived from the model by principal component analysis, which is the result of the classification. The inferred model is based on the chemical composition of each glass type, and the error tolerance of the glass type is 0.18, which is reasonable. The regression equation of the principal component analysis did not change the glass type when the data was perturbed by ±5% for silica, which has a large influence on the principal component. The results do not change when the data are perturbed.

In identifying the glass type, the inferred equation for each glass type was obtained by using principal component analysis, and the distance to the center of each glass number was obtained by combining the data with cluster analysis. The sensitivity analysis of the data perturbation is also based on the robustness. When the data are perturbed, the results do not changes.

References

[1] Ji Luoyuan, PJ Cherian. Peacock blue-glazed pottery specimens excavated from Kerala, India [J]. Journal of the Palace Museum, 2022(06):55-67+148.DOI:10.16319/j.cnki.0452-7402.2022.06.004.

[2] Pan Ling, Tan Wenyu. The cultural factors of the west in the Xianbei relics of Hulunbuir--and the "Grassland Silk Road" in the Han Dynasty [J]. Archaeology,2022(05):110-120+2.

[3] Huang Qiaohao. From the local to the exotic, experiencing the "lucidity" of sea silk glassware [J]. Collection. Auction, 2021(06):56-61.

[4] Wu Yongmei. Witnesses to the civilization of the Silk Road in the Middle Ages: wine and its usage in the Tang Dynasty [J]. Oriental Collection,2021(19):97-98.

[5] Di Jin, Zhang Wenli, Du Ying. Tracing the "lucite" across a thousand years [N]. Science and Technology Daily,2021-09-03(008).DOI:10.28502/n.cnki.nkjrb.2021.004953.

[6] Hirayama Ikuo. Silk Road Museum Collection of Ancient Glassware Exhibited at Dunhuang [J]. Heritage Identification and Appreciation, 2021(15):102.

[7] Xu Siwen, Gu Zhou, Yang Yimin. The distribution of Indo-Pacific beads and the Maritime Silk Road before the 2nd century A.D. [J]. Journal of Chinese History of Science and Technology, 2021, 42(02):281-292+321.

[8] Wang S. Y. Maritime silk road relics of the Han dynasty in Lingnan from the exhibition of "Silk Road Voyage"[J]. Collection, 2021(02):162-171.

[9] Zhou Tingxi. From the Mediterranean to China: An Exploration of the Collection of Cultural Objects in the Silk Road Museum of Art by Ikuo Hirayama [J]. Artworks,2020(12):44-49.

[10] Zhang X M , Feng Q H , Wang X S , et al. Establishment and Application of Material Balance Equations for Low-rank Coalbed Methane Reservoirs[J]. Natural Gas Geoscience, 2013, 24(6):1311-1315.