Chemical composition correlation between high potassium glass and lead-barium glass

Zirong Zhang*, Fangyuan Zhu, Runhao Chen

School of Electro-mechanical Engineering, Guangdong University of Technology, Guangzhou, China

*Corresponding author: zzr13433782589@163.com

Abstract: The Silk Road was a channel for cultural exchange between China and the West in ancient times, of which glass was a valuable physical evidence of early trade exchanges. Early glass was often made into bead-shaped ornaments in the West Asian and Egyptian regions into our country, and our ancient glass was made locally after absorbing its technology, so it was similar in appearance to foreign glass products, but the chemical composition was not the same. As glass is highly susceptible to weathering by the burial environment, during the weathering process, a large number of internal elements are exchanged with environmental elements, resulting in a change in its composition ratio, which affects the correct judgment of its category. According to the proportion of chemical composition of glass artifacts, it was split into high potassium glass data and lead-barium glass data, and the Pearson correlation coefficient was studied by using matlab software for high potassium glass data and lead-barium glass data respectively, so as to analyze the correlation between each chemical composition.

Keywords: High potassium glass; lead-barium glass; Pearson correlation coefficient; gray correlation analysis.

1. Introduction

The Silk Road was a channel for cultural exchange between China and the West in ancient times, of which glass was a valuable physical evidence of early trade exchanges[1]. Early glass was often made into bead-shaped jewelry in the West Asian and Egyptian regions into our country, our ancient glass absorbed its technology and made in situ locally, so the appearance of glass products with foreign similarities, but the chemical composition is not the same[2].

The main raw material of glass is quartz sand, the main chemical composition is silicon dioxide (SiO2). Due to the high melting point of pure quartz sand, in order to reduce the melting temperature, it is necessary to add fluxes when refining[3]. The fluxes commonly used in ancient times were grass ash, natural alkali, saltpeter and lead ore, and limestone was added as a stabilizer, which was converted to calcium oxide (CaO) after calcination[4]. Adding different fluxes, its main chemical composition is also different. For example, lead-barium glass adds lead ore as a flux in the firing process, and its content of lead oxide (PbO) and barium oxide (BaO) is higher, which is usually considered to be our own invented glass species, and the glass of Chu culture is mainly lead-barium glass. Potassium glass is fired with substances containing high levels of potassium such as grass wood ash as a flux, and is mainly popular in Lingnan, China, as well as in Southeast Asia and India and other regions[5].

Ancient glass is highly susceptible to weathering by the burial environment. During the weathering process, internal elements were exchanged with environmental elements in large quantities, resulting in changes in the proportion of its composition, which affects the correct judgment of its category[6].

There is a batch of data related to ancient glass products in China, and archaeologists have classified them into two types of high potassium glass and lead-barium glass based on the chemical composition of these artifact samples and other testing methods. In this paper, we analyze the correlations between the chemical compositions of glass artifact samples of different categories and compare the differences of the correlations between the chemical compositions of different categories.

2. Model assumptions and notation

2.1 Assumptions[7]

1. The data are real and reliable, and representative.
2. No consideration of errors in data processing and graph drawing by software tools.
3. For the time being, chemical components other than those selected in this paper are not considered.
4. Assume that the data involved in the model are all original and valid, without considering the existence of special cases.
5. Assume that the effect of time is not considered when considering the changes in glass type as well as chemical composition content before and after weathering.

2.2 Notations

Important notations used in this paper are listed in Table 1.

Table 1 Notations

Symbols	Description
$x_1, x_2, \cdots, x_p$	Each chemical component variable
R	Correlation coefficient matrix
μ_i	Clustering center
$\mathrm{J}(C)$	Distance sum of squares
T	Number of leaf nodes of decision tree
γ	Penalty term for the number of leaf nodes
ω_j	Score of the jth leaf node
λ	L_2 canonical penalty term
K	Number of decision trees
w_i	Weights

3. Model construction and solving

3.1 Correlation between the chemical composition of different categories of glass artifact samples

With random variables X and Y for which n random trials are conducted, the observations obtained are $(X_i, Y_i)(i = 1,2, \ldots, n)$, $\bar{X}$, $\bar{Y}$ are the respective expected values, $\sqrt{DX}$ and $\sqrt{DY}$ are the variances of *X* and *Y*, respectively. *Cov(X,Y)* is the covariance, r is the correlation coefficient, and *R* is the correlation coefficient between the random variables *X* and *Y* for the sample $(X_i, Y_i)(i = 1,2, \ldots, n)$, which is called the sample correlation coefficient. For a relationship with multiple variables, where the random variables *X* and *Y* can represent any two of the random variables[8]. In practice, the sample correlation coefficient *R* is often used as an estimate of the correlation coefficient r.

Based on the above theoretical knowledge the following correlation coefficient model can be developed.

$$\begin{cases} r = \frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}} \\ R = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2 \sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \\ Cov(X,Y) = E[(X-\bar{X})(Y-\bar{Y})] \\ \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \\ \bar{Y}\,\frac{1}{n}\sum_{i=1}^{n} Y_i \end{cases} \tag{1}$$

According to the principle of multivariate statistical analysis, the Pearson correlation coefficient r_{XY} of two random vectors X and *Y* can be obtained by the following equation.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(Y)^2}{N}\right)}} \quad (2)$$

where N is the number of samples and all summations are from 1 to N. The value of the correlation coefficient r is between -1 and +1, i.e. $-1 \leq r \leq +1$. The properties are as follows.

1) when $r > 0$, it means that the two variables are positively correlated, and when $r < 0$, the two variables are negatively correlated.

2) when $|r|=1$, it means that the two variables are fully linearly correlated, i.e., a functional relationship.

3) when $r=0$, it means that there is no linear correlation between the two variables.

4) when $0<|r|<1$, it means that there is some degree of linear correlation between the two variables. And the closer $|r|$ is to 1, the closer the linear relationship between the two variables; $|r|$ is closer to 0, the weaker the linear correlation between the two variables.

It can be generally classified by three levels.

A. $|r|<0.4$ for low linear correlation.

B. $0.4 \leq |r| < 0.7$ as significant correlation.

C. $0.7 \leq |r| < 1$ for high linear correlation.

According to the proportions of chemical composition of all glass artifacts collected, they need to be split into high potassium glass data and lead-barium glass data, so that the correlations between chemical compositions can be analyzed for different categories of glass artifact samples.

3.1.1 Correlation of the composition of high potassium glass

Table 2 Correlation coefficient matrix

R	SiO_2	Na_2O	K_2O	SrO	SnO_2	CaO
SiO_2	1.000***	-0.475**	-0.781***	-0.524**	0.070	-0.752***
Na_2O	-0.475**	1.000***	0.606***	-0.110	-0.108	0.640***
K_2O	-0.781***	0.606***	1.000***	0.404*	0.070	0.690***
SrO	-0.524**	-0.110	0.404*	1.000***	0.306	-0.052
SnO_2	0.070	-0.108	0.070	0.306	1.000***	-0.374
CaO	-0.752***	0.640***	0.690***	-0.052	-0.374	1.000***

Note: ***, **, * represent 1%, 5%, 10% significance levels, respectively

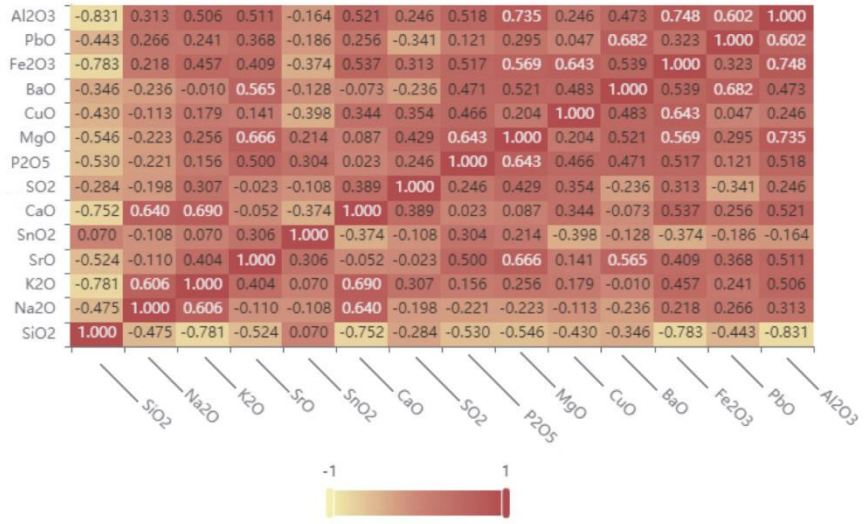


Figure 1 Correlation coefficient heat map

The Pearson correlation coefficients of the high potassium glass data were investigated by using the corrcoef function (set significance level $\alpha = 0.05$) in matlab software, and the Pearson correlation coefficient matrices of the 14 chemical components were obtained partially as shown in Table 2, and the complete table is shown in the Appendix, and the correlation coefficient heat map is shown in Figure 1.

From the knowledge of statistics, it is known that the closer the correlation coefficient is to ±1, the stronger the correlation between the two variables, i.e., when one of the variables changes, the other variable will change with it; while the closer the correlation coefficient is to 0, the more independent the two variables are, i.e., when one of the variables changes, the other variable will not change at will. From the above graph, we can see that the darker the color the stronger the correlation.

3.1.2 Correlation of the composition of lead barium glass

The Pearson correlation coefficient of lead-barium glass data was studied by using the corrcoef function (set significance level $\alpha=0.05$) in matlab software, and the Pearson correlation coefficient matrix of 14 chemical components was partially obtained as shown inTable 3, and the correlation coefficient heat map is shown in Figure 2.

Table 3 Correlation coefficient matrix

R	SiO_2	Na_2O	K_2O	SnO_2	SrO	CaO
SiO_2	1.000***	0.362**	0.087	0.085	-0.503***	-0.488***
Na_2O	0.362**	1.000***	-0.059	-0.083	-0.093	-0.373***
K_2O	0.087	-0.059	1.000***	0.195	-0.121	0.096
SrO	0.085	-0.083	0.195	1.000***	-0.038	0.210
SnO_2	-0.503***	-0.093	-0.121	-0.038	1.000***	0.171
CaO	-0.488***	-0.373***	0.096	0.210	0.171	1.000***

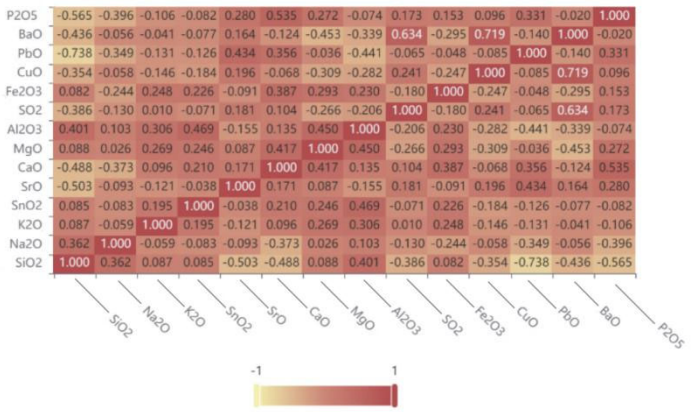


Figure 2 Correlation coefficient heat map

From the knowledge of statistics, it is known that the closer the correlation coefficient is to ±1, the stronger the correlation between the two variables, i.e., when one of the variables changes, the other variable will change with it; while the closer the correlation coefficient is to 0, the more independent the two variables are, i.e., when one of the variables changes, the other variable will not change at will. From the above graph, we can see that the darker the color the stronger the correlation.

3.2 Variance Analysis

High potassium glass is mainly influenced by silica (SiO2) and calcium oxide (CaO), and lead barium glass is mainly influenced by barium oxide (BaO), silica (SiO2) and lead oxide (PbO), then the silica (SiO2) components that are both available can be selected to compare the differences in the chemical composition correlations between different categories. That is, the silica (SiO2) chemical composition is selected as the dependent variable (parent series) for the analysis, and the remaining appropriate variables are selected as independent variables (subsequences), and a gray correlation analysis model is established to calculate the case of its gray correlation, and the gray correlation calculated in different categories of glass is subjected to ANOVA to observe the variability between high potassium glass and lead-barium glass.

Gray system theory is an emerging theory proposed by Professor Deng Julong in 1982, which is a new system theory using a specific method to describe the system with incomplete information and make prediction, decision and control[9,10].

Gray correlation analysis is a new multi-factor analysis method, whose basic principle is to distinguish the closeness of the relationship between multiple factors in a system by comparing the geometric relationship of statistical series, the closer the geometric shape of the series curves, the greater the gray correlation between them, and vice versa, the smaller the correlation. In the system analysis of complex systems, most of them used statistical analysis or other analysis methods in the past. The advantages of gray correlation analysis over them are: gray correlation analysis is done according to the development trend, so there is no excessive requirement for the amount of sample size, and there is no need for typical distribution laws, and the calculation is relatively small, and the results will be more consistent with the results of qualitative analysis. Therefore, gray correlation analysis is a relatively simple and reliable analysis method in system analysis. The specific calculation steps of gray system correlation analysis are as follows.

1.Determination of the reference series and comparison series

(1) Reference series - the data series reflecting the characteristics of the system behavior

(2) Comparison series - the data series of factors affecting the behavior of the system

2.Quantumless processing of reference and comparison series

(1) initialization - each number in the matrix is divided by the first number to get a new matrix

(2) mean value - each number in the matrix is divided by the average of all elements of the matrix to obtain a new matrix

3. Find the gray correlation coefficient $\xi(X_i)$ of the reference and comparison series

Let the reference series be X_0, the comparison series be X_1,X_2,X_3..., and the correlation coefficient $\xi(X_i)$ of the comparison series with respect to the reference series at each point of the curve

$$\xi(X_i) = \frac{\min\min|\hat{X}^{(0)}(k)-X^{(0)}(k)|+\rho\max\max|\hat{X}^{(0)}(k)-X^{(0)}(k)|}{|\hat{X}^{(0)}(k)-X^{(0)}(k)|+\rho\max\max|\hat{X}^{(0)}(k)-X^{(0)}(k)|} \tag{3}$$

Correlation coefficient It can also be simplified as the following equation

$$\xi_{0i} = \frac{\Delta(min)+\rho\Delta(max)}{\Delta_{0i}(k)+\rho\Delta(max)} \tag{4}$$

4.Find the correlation degree r_i

The correlation coefficient - the value of the degree of correlation between the comparison series and the reference series at each moment (i.e., each point in the curve), so it has more than one number, and the information is too scattered to facilitate a holistic comparison. Therefore it is necessary to pool the correlation coefficients at each moment (i.e., at each point in the curve) into one value, i.e., to find its average value, as a quantitative representation of the degree of correlation between the comparison series and the reference series. The formula for the degree of correlation between $X^{(0)}(k)$ and $\hat{X}^{(0)}(k)$ is as follows.

$$r = \frac{1}{n}\sum_{k=1}^{n}\eta(k) \tag{5}$$

5. Order of correlation

The degree of correlation between factors is mainly described by the order of the magnitude of the correlation, not only the size of the correlation. The correlation order is formed by arranging the correlation degree of m sub-sequences to the same parent sequence in order of size, which is called {x}, and it reflects the "superiority" relationship of each sub-sequence to the parent sequence. If $r_{0i} > r_{0j}$, then $\{x_i\}$ is said to be better than $\{x_j\}$ for the same parent sequence $\{x_0\}$, denoted as $\{x_i\} > \{x_j\}$.

(CuO), iron oxide (Fe2O3), lead oxide (PbO), barium oxide (BaO), sulfur dioxide (SO2), tin oxide (SnO2), phosphorus pentoxide (P2O5), strontium oxide (SrO) and silicon dioxide (SiO2), and provide an analytical reference based on the correlation degree, and the discrimination coefficient is taken as 0.5 when using gray correlation analysis. The correlation coefficients are calculated. The correlation coefficient of high potassium glass, for example, is shown in Figure 3 below.

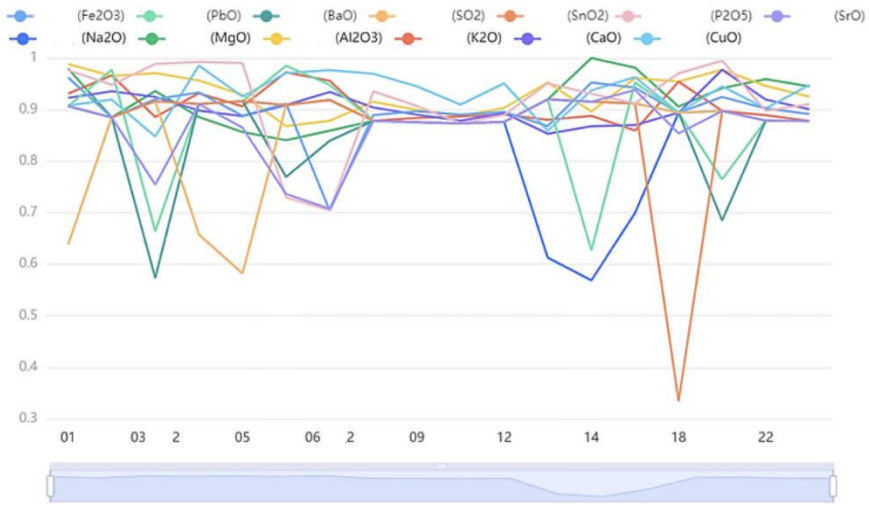


Figure 3 Correlation coefficient chart

Combined with the correlation coefficient calculation formula to calculate the correlation coefficient value, and according to the correlation coefficient value, and then calculate the correlation degree value for evaluation judgment, the calculated gray correlation degree is shown in Table 4 below.

Table 4 Gray correlation

Evaluation item	High potassium glass correlation	Lead barium glass correlation
Aluminum oxide(Al2O3)	0.932	0.953
Copper oxide(CuO)	0.931	0.915
Phosphorus pentoxide(P2O5)	0.917	0.896
Magnesium oxide(MgO)	0.911	0.931
Potassium oxide(K2O)	0.907	0.92
Calcium oxide(CaO)	0.903	0.921
Iron oxide(Fe2O3)	0.897	0.911
Lead oxide(PbO)	0.873	0.934
Tin oxide(SnO2)	0.866	0.884
Strontium oxide(SrO)	0.864	0.924
Barium oxide(BaO)	0.855	0.936
Sulfur dioxide(SO2)	0.85	0.882
Sodium oxide(Na2O)	0.849	0.9

The correlation coefficients were combined with the above correlation coefficients and weighted to obtain the correlation value, which was used to rank the 13 evaluation items; the correlation value ranged from 0 to 1, and the larger the value was, the stronger the correlation with the "reference value" (parent series), which means the higher the evaluation. As can be seen from the above table, for the 13 evaluated items, aluminum oxide (Al2O3) was the most highly evaluated (correlation: 0.932), followed by copper oxide (CuO) (correlation: 0.931). The results of the variance analysis using the correlations are shown in Table 5 below.

Table 5 Variance Analysis

Variable name	Median	Mean value	Standard deviation	Skewness	Kurtosis	S-W test
High potassium glass correlation	0.897	0.889	0.03	0.021	-1.616	0.906(0.162)
Lead-barium glass correlation	0.92	0.916	0.021	-0.155	-0.514	0.969(0.877)

4. Conclusion

In this paper, analyzed the correlations between the chemical compositions of glass artifact samples of different categories and compared the differences of the chemical composition correlations between the different categories. They were split into high potassium glass data and lead-barium glass data, and Pearson correlation coefficients were studied for high potassium glass data and lead-barium glass data respectively using matlab software, so as to analyze the correlations between chemical compositions. Next, the chemical composition of silica (SiO2) was selected as the dependent variable (parent series) for the analysis, and the remaining variables were selected as independent variables (subseries), and a gray correlation analysis model was established to calculate its gray correlation, and the calculated gray correlations in different categories of glass were subjected to ANOVA to observe the differences in the correlations between the chemical compositions of high-potassium glass and lead-barium glass.

References

[1] WEN Rui,ZHAO Zhiqiang,MA Jian. Compositional analysis of glass beads excavated from the Shirenzigou site group in Balikun, Xinjiang[J]. Spectroscopy and Spectral Analysis,2016,36(09):2961-2965.

[2] Cheng Qian,Guo Jinlong,Wang Bo.LA-ICP-AES analysis of the compositional characteristics of glassware excavated from the Silk Road and the end of the Silk Road[J]. Spectroscopy and Spectral Analysis,2012,32(07):1955-1960.

[3] XU Zhenqi, ZHAO Hongxia, KAN Fuxi. Nondestructive analysis of chemical composition, structure and physical phases of natural glasses[J]. Journal of Silicates,2012,40(03):443-449.

[4] Cui Jianfeng,Wu Xiaohong,Tan Yuanhui. Composition analysis of ancient glassware excavated from Chu tombs of the Warring States period in the Yuanshui basin, Hunan[J]. Journal of Silicate,2009,37(11):1909-1913+1918.
[5] Fu Xiu-Feng, Gan Fu-Xi. Study on the composition of a group of ancient glasses from South and Southwest China based on multivariate statistical analysis[J]. Conservation and Archaeological Science, 2006(04):6-13.
[6] Liu Zhuang, Zhang Zhihong. Analysis of the composition of glass beads excavated from the Dashi site of Northern Wei in Luoyang[J]. Spectroscopy Laboratory, 1999(01):75-79.
[7] Yang Liangliang. K-means clustering for subtle features in radio station sorting identification technology[J]. Telecommunications Technology,2022,62(08):1149-1154.
[8] Han L, Xu Qiao. Construction of a model for the effect of dextrorphan for preterm abortion based on XGBoost algorithm[J]. Chinese Sex Science,2022,31(08):117-121.
[9] Zhou G.C., Meng Q.H., Bao Chunyu. Gray correlation analysis of lower limb kinematic factors affecting the jump height of volleyball players[C]//. Abstracts of the 22nd National Conference on Sports Biomechanics,2022:558.
[10] Zhou NNA, Rao CJ. Analysis of factors influencing grain yield in Yunnan Province based on gray correlation analysis[J]. Agriculture and Technology,2022,42(15):164-167.