Summary of the methods of composition analysis and identification of ancient glass objects

Ziyang Tang, Jiayi Zhao*

SDU-ANU Joint Science College, Shandong University, Shandong, WeiHai, 264209

*Corresponding author: 202000700259@mail.sdu.edu.cn

Abstract. The analysis and identification of the composition of ancient glass objects is a major topic in archaeological research today, and its main task is to restore the true appearance of ancient artifacts as completely as possible by examining various characteristics (surface weathering, color, chemical composition, etc.). The analysis of the indicators before and after weathering at the sampling points is a very important research method that should be continuously improved and rationalized. In this thesis, the characteristics of various types of glass products before and after weathering and the indicators of chemical composition are analyzed qualitatively and quantitatively through various methods (modeling, data analysis, etc.), and the patterns among the indicators of various types of glass are identified, and then the composition of various types of glass before weathering is predicted and the type of glass of unknown glass artifacts is judged. The ideas and research methods embodied therein are of certain significance for the analysis and identification of ancient glass objects in the future.

Keywords: principal component analysis, cluster analysis, sensitivity analysis.

1. Introduction

The analysis and research on the composition of ancient glass in China began in the middle of the last century, because the materials of ancient glass in China originated from the local area and learned from the western technology, so although the appearance of glass products is similar to that of foreign ones, there are differences in the types and contents of the chemical components that make up the glass, and Chinese ancient glass is characterized by the coexistence of high PbO and BaO, mainly potassium silicate glass (with substances containing high potassium such as oxalic ash as a flux), soda-calcium silicate, alkali-containing calcium-silicate glass, and lead-barium glass (with lead ore added as a flux in firing), and other categories.[1] First, we classify the glasses separately into broad categories and subcategories according to the measured classification criteria. The sensitivity was analyzed by means of a constructed function. The low sensitivity of our classification model was illustrated by the way that the value of the effect on the function was found to be smaller by changing a component variable. Based on the screening data, we selected the more important compounds in the glass and the correlation analysis between the principal components. Scatter plots were first drawn to determine whether the two had a linear relationship. After judging that none of them had linear relationship, then we analyzed the correlation between compounds using Spearman's correlation coefficient and summarized the patterns and differences.

2. Methods

In this paper, we used correlation analysis, principal component analysis and cluster analysis to solve the problem of analyzing and identifying the components of ancient glass products while conducting mathematical modeling.

(1) Identification of glass types with known chemical compositions, as well as analysis of the rationality of the identification results. It is first necessary to classify the glass into major classes and subclasses respectively according to the classification criteria specified in the first two questions. Then a sensitivity analysis system is established, and here we can use the construction of a sensitivity function to analyze the sensitivity. Assuming a small change in the independent variable, observe

whether the function will change significantly as a way to determine the sensitivity of our classification model.

(2) Analyze the correlation of chemical components in each category of glass and compare the variability of chemical component correlations between different categories. Firstly, we can keep the relatively important chemical components in the glass and determine whether they have a linear relationship between two of them. Then the correlation coefficients are calculated for each of them, and the tables are checked to see whether they are significantly correlated. Finally, the variability of chemical components between different categories is compared.

3. Model assumptions

(1) There is no correlation between the data of each glass artifact sample, i.e., the data of different types of glass are independent of each other.

(2) The given measurement data are accurate and there is no error in judging the glass types.

(3) The chemical composition of the artifact samples was not affected by external environmental factors other than weathering.

(4) the glass type is considered only with or without weathering of high potassium glass, lead-barium glass.

(5) the effect of oxidation on the surface and internal oxide composition of the glass is not considered.

4. Model building and solution

4.1. Classification of high potassium and lead-barium glass subclasses

4.1.1 Classification of subclasses of lead-barium glass

For the subclass division of high potassium and lead-barium glass, considering that if all 14 chemical components are involved in the principal component algorithm, the obtained principal component results cannot meet the needs of the actual situation, therefore, according to the content of each chemical component to determine its importance, six chemical components (SiO_2, CuO, PbO, BaO, Al_2O_3, P_2O_5) are selected to participate in the principal component analysis, and their contents are .

The principle of principal component analysis is described below[2].

Let the number of samples be , and the number of indicators be (in the problem), which can form the sample matrix

$$\mathbf{x}=\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}=(x_1,\cdots\cdots,x_n). \tag{1}$$

Afterwards, the data is normalized: (here is the label of the data in the matrix, which is different from the meaning above)

Average value: $\overline{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$, Standard deviation $s_j=\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2\right]^{\frac{1}{2}}$,

thus obtaining standardized data $X_{ij}=\frac{x_{ij}-\overline{x}_j}{s_j}$,

Normalization matrix of the sample

$$\mathbf{X}=\begin{pmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{m1} & \cdots & X_{mn} \end{pmatrix}=(X_1,\cdots\cdots,X_n) \tag{2}$$

The covariance matrix of the standardized sample is calculated

$$\mathbf{R}=\begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{pmatrix} \tag{3}$$

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

Calculate the eigenvalues and eigenvectors of The eigenvalues , and the corresponding eigenvectors are

$$a_1=\begin{pmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{n1} \end{pmatrix},\quad a_2=\begin{pmatrix} a_{12} \\ a_{22} \\ \cdots \\ a_{n2} \end{pmatrix},\quad \cdots\cdots,\quad a_n=\begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdots \\ a_{nn} \end{pmatrix} \tag{4}$$

Principal component contribution rate $\mu_i=\frac{\lambda_i}{\sum_{k=1}^{n}\lambda_k}$, $i=1,2,\cdots\cdots,p$, Cumulative contribution rate $\upsilon_i=\frac{\sum_{k=1}^{i}\lambda_k}{\sum_{k=1}^{n}\lambda_k}$.

Generally, the first principal component corresponding to the eigenvalue of the cumulative contribution is taken, where the first principal component

$$F_i=a_{1i}X_1+a_{2i}X_2+\cdots\cdots+a_{ni}X_n,\ i=1,2,\cdots\cdots,t \tag{5}$$

For a given principal component, the larger the coefficient in front of the indicator, the greater the influence of the indicator on that principal component is proved.

Applying the above principle, the following results can be obtained.

Table 1 MATLAB principal component analysis results

Accumulative Contribution Rate	0.4384	0. 6673	0.8468	0.9734	0.9957	1.0000
Eigenvector matrix (Percentage of components)	-0.5327	-0.0982	0.3396	0.3649	-0.1681	-0.6556
	0.4904	-0.3694	0.3181	0.0486	-0.7200	0.0334
	0.2735	0.7142	-0.2201	-0.2039	-0.3132	-0.4764
	0.5550	-0.2727	0.1756	-0.0808	0.5665	-0.5093
	-0.2647	-0.4591	-0.3807	-0.6927	-0.1715	-0.2549
	0.1513	-0.2423	-0.7478	0.5801	-0.0700	-0.1331

As shown in Table 1, the first three main components were identified as:

$$F_1=-0.5327p(x_0)+0.4904p(x_7)+0.2735p(x_8)+0.555p(x_9)-0.2647p(x_5)+0.1513p(x_{10})$$

$$F_2=-0.0982p(x_0)-0.3694p(x_7)+0.7142p(x_8)-0.2727p(x_9)-0.4591p(x_5)-0.2423p(x_{10})$$

$$F_3=0.3396p(x_0)+0.3181p(x_7)-0.2201p(x_8)+0.1756p(x_9)-0.3807p(x_5)-0.7478p(x_{10})$$

Through the results of principal component analysis, we can analyze that

Principal component I consists of SiO_2, BaO and CuO with the largest composition ratio, and it is known from the main composition of glass and the firing process of lead-barium glass that principal component II mainly describes the most important compound composition of lead-barium. SiO_2 carries a larger negative load, while BaO and CuO carry a relatively small positive load, so principal component I can be used to reflect the proportion of flux added in firing; principal Component II is mainly composed of PbO and Al_2O_3, and component III is mainly composed of P_2O_5 and Al_2O_3, both of which together reflect the weathering characteristics of lead-barium glass[3].

After dimensionality reduction of the data, we can use K-means clustering to divide the subclasses. It is worth noting that here, unlike the previous subquestion, we do not know exactly how many classes we need to divide, so we need to use the elbow diagram to help us see how many classes we need to divide to get a reasonable classification effect.

The calculated values of the three principal components of each artifact were imported into spss for cluster analysis. The elbow diagram for the classification of lead-barium glass according to the spectrogram is as in the Figure 1 shown.

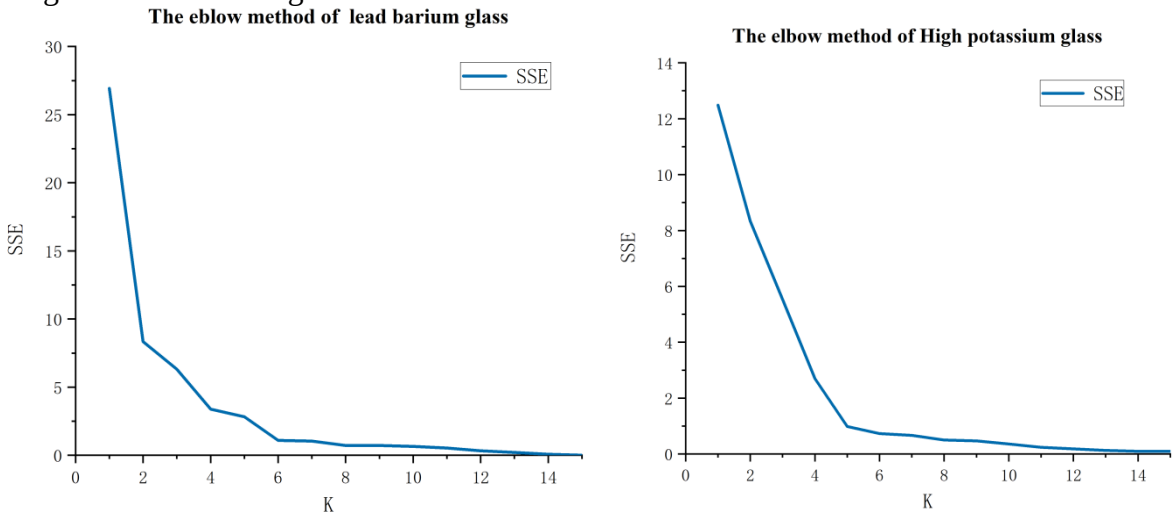


Figure 1 Elbow diagram of the cluster analysis of lead-barium and high potassium

After extracting the coefficients of the system clustering, the elbow diagram of the coefficients is drawn in descending order, and according to the analysis for the elbow diagram, a relatively small slope is chosen as a reasonable number of classifications, which is 4 classes.

The classifications are introduced as follows.

(1) SiO_2-CuO-PbO-BaO glass, which contains more CuO in addition to more SiO_2 and flux, but there is only one class among all undifferentiated glasses.

(2) SiO_2-PbO-BaO-Al_2O_3 glass, in addition to containing more SiO_2 and combustion aids, but also contains more Al_2O_3, but also the same only one class.

(3) high melting point SiO2-PbO-BaO glass, which contains less PbO and BaO, as PbO and BaO are converted from flux, the glass may have a high melting point.

(4) Low melting point SiO_2-PbO glass, in which the PbO and BaO content is higher, and since PbO and BaO are converted from fluxes, this glass may have a lower melting point. The latter two occupy a larger number of species.

4.1.2 Classification of high potassium glass subclasses

Similar to lead-barium glass, for high potassium glass, according to the problem a content analysis of compounds in high potassium glass, after removing the compounds with low content and relatively stable composition in high potassium glass, five compounds of SiO_2, Na_2O, K_2O, CaO and Al_2O_3 were retained, and they were subjected to principal component analysis, and two principal components were chosen to be retained.

The two principal components were analyzed as follows.

Principal component one consists of SiO_2, K_2O and also CaO, the main components of high potassium glass, where both K_2O and CaO have large positive loadings, while SiO_2 carries a large negative loading. Since the three compounds occupy the main composition of the high potassium glass and are mainly self-contained elements of the glass itself, composition I describes the basic compound composition of the high potassium glass; the main composition II consists of Na_2O, Al_2O_3, and Na_2O with a very high negative load, while alumina with a small positive load. The above three compounds are mainly derived from weathering, so principal component II mainly describes the weathering condition of the high-potassium glass.

The data obtained from the principal component analysis were then imported into spss for systematic cluster analysis, and the spectrograms are shown in the Appendix. According to the spectral diagram, combined with the elbow diagram and the complexity of the classification, the high-potassium glasses were divided into two subclasses.

(1) ordinary high potassium glasses

(2) High potassium glasses with the presence of $CaO\text{-}K_2O\text{-}Al_2O_3$. These glasses contain not only the compounds found in ordinary high potassium glasses, but also CaO, K_2O, and Al_2O_3, which are presumed to be elements added during refining to enhance the color and properties of the glass.

In general, the results we obtained from data dimensionality reduction by correlation analysis and principal component analysis, and then subtype classification by clustering method, are consistent with the classification results of literature data to some extent. We can say that our division is very reasonable.

4.2. Classification prediction of high-potassium and lead-barium glass

In this chapter, we will identify the glass types with known chemical composition, as well as analyze the reasonableness of the identification results. Our solution is divided into the following two steps.

(1) Classify the glass into major classes and subclasses according to the classification criteria specified in the first two questions, respectively.

(2) Analyze the sensitivity by means of a construction function.

The determination methods are as follows.

Method (1): It is easy to analyze the data: high potassium glass before weathering has K_2O (x_2) as the main component in addition to SiO_2, and lead-barium glass before weathering has PbO (x_8) and BaO (x_9) as the main components in addition to SiO_2. Therefore, the elemental content of the two types of glass before weathering can be considered as the basis for judgment. If satisfied

$$2q(x_2) < q(x_8) + q(x_9) \tag{6}$$

Then the glass type is lead-barium glass, otherwise it is high potassium glass, where

$$q(x_n) = \frac{p(x_n)}{100 - p(x_0)},\ (n=1,2,\cdots\cdots,13) \tag{7}$$

is the percentage content of the remaining components with SiO_2 removed.

After weathering, the K_2O content of high potassium glass is greatly reduced, so it can no longer be judged by the above method.

Method (2): Since the content of PbO and BaO in the glass remains basically the same before and after weathering, if the content of PbO and BaO in the glass is extremely low, regardless of before and after weathering, the glass is most likely high potassium glass.

Method (3): The SiO_2 content remains basically unchanged (or even increases) before and after weathering for high potassium glass, and the SiO_2 content is severely reduced after weathering for lead-barium glass.

Method (4): The SO_2 content of severely weathered lead-barium glass is high, while the SO_2 content of high potassium glass does not change much before and after weathering.

However, it is necessary to find a scientific and effective method, and the concept of "characteristic point" is introduced below.

Known type of glass j ($j=1,2,3,4$), corresponding to: (high potassium unweathered, high potassium weathered, lead-barium unweathered, and lead-barium weathered) four glass types, respectively. Using the percentages of the remaining components excluding SiO_2 defined in problem one (2) $q(x_n)$, Let the nth chemical composition of the jth glass be x_{jn}, $S=[0,1]$.

In space S^{13}, point $q(x^j)=(q(x_1^j),q(x_2^j),\cdots\cdots,q(x_{13}^j))$, the meaning of point $q(x^j)$ is the content of each chemical component in the data. The goal is to find $q(x_j)=(q(x_{j1}),q(x_{j2}),\cdots\cdots,q(x_{j,13}))$ makes

$$\begin{aligned}&\min\ z=\sum_{l=1}^{m}\left[\sum_{n=1}^{13}(q(x_n^j)-q(x_{jn}))^2\right]^{\frac{1}{2}},\ j=1,2,3,4\\&s.t.\quad q(x_n^j)\in[0,1]\end{aligned}\tag{8}$$

established.

The geometric meaning is: find the point $q(x^j)=(q(x_1^j),q(x_2^j),\cdots\cdots,q(x_{13}^j))$ such that the sum of its distances from the remaining m points (the points formed by the data in the table) is minimized.

The point $q(x^j)$ thus obtained is defined as the "characteristic point" of the jth glass, i.e., the point that best reflects the chemical composition of the jth glass.

Based on the "feature points", we can obtain the following quantitative calculation method.
Method (5).

Set the composition points of each chemical composition of the unknown glass type , find the suitable glass type , so that

$$\min\ d=\left[\sum_{n=1}^{13}(q(x_n)-q(x_n^j))^2\right]^{\frac{1}{2}},\ j=1,2,3,4\tag{9}$$

established.

Judgment is made using the above problem classification method. Judgment results.

Table 2 Prediction results for glass types in Data Form III

Glass	Broad category division	Subclass division
A1	High Potassium Glass	Ordinary high potassium glass
A2	Lead barium glass	Low melting point SiO_2-PbO glass
A3	Lead barium glass	High melting point SiO_2-PbO-BaO glass
A4	Lead barium glass	SiO_2-PbO-BaO-Al_2O_3 glass
A5	High Potassium Glass	CaO-K_2O-Al_2O_3 High Potassium Glass
A6	High Potassium Glass	Ordinary high potassium glass
A7	High Potassium Glass	Ordinary high potassium glass
A8	Lead barium glass	SiO_2-CuO-PbO-BaO glass

4.3. Reasonableness and sensitivity

4.3.1 The classification results and the explanation of reasonableness are now as follows.

(a) A1 has low content of all compounds except K_2O and SiO_2, and belongs to ordinary high potassium glass.

Only the content of PbO and SiO_2 is high in A2, which belongs to low melting point SiO2-PbO glass.

A3, in which both PbO and BaO are high, belongs to high melting point SiO_2-PbO-BaO glass.

A4 in addition to the characteristics of high melting point SiO_2-PbO-BaO glass, the Al_2O_3 content is also high, so it belongs to SiO_2-PbO-BaO-Al_2O_3 glass.

A5 in which the contents of CaO, K_2O and Al_2O_3 are all higher, belongs to CaO-K_2O-Al_2O_3 high potassium glass.

A6, A7, all with SiO_2 content higher than 90 percent, belong to ordinary high potassium glass.

The content of SiO_2, CuO, PbO and BaO in A8 is higher, so it belongs to SiO_2-CuO-PbO-BaO glass.

4.3.2 Sensitivity analysis.

Using the method (5) mentioned to solve problem, the method uses the function

$$d=\left[\sum_{n=1}^{13}(q(x_n)-q(x_n^j))^2\right]^{\frac{1}{2}},\ j=1,2,3,4 \tag{10}$$

The essence of the judgment is to follow the least squares method in linear regression fitting, in which the data of unknown artifacts are fitted against the known data of various glass types, and the type of glass is corresponded to the good or bad effect of the fitting, and finally a conclusion is drawn. The method of thought used is an innovation and application of the classical method, and therefore possesses a certain degree of rationality.

The sensitivity of the method is analyzed below.

Discuss the effect of a small change in the independent variable $q(x_n)$ on the dependent variable d.

The sensitivity of the dependent variable to the independent variable is measured by the relative change measure, and the sensitivity of d to $q(x_n)$ is noted as $T(d,q(x_n))$ and defined as

$$T(d,q(x_n))=\frac{\Delta d/d}{\Delta q(x_n)/q(x_n)}\approx\frac{d}{dq(x_n)}(d)\frac{q(x_n)}{d} \tag{11}$$

After calculation, we get

$$\frac{d}{dq(x_n)}(d)=\frac{q(x_n)-q(x_n^j)}{\left[\sum_{n=1}^{13}(q(x_n)-q(x_n^j))^2\right]^{\frac{1}{2}}} \tag{12}$$

So get

$$T(d,q(x_n))=\frac{q(x_n)[q(x_n)-q(x_n^j)]}{\sum_{n=1}^{13}(q(x_n)-q(x_n^j))^2} \tag{13}$$

Since by the definition of $q(x_n)$, $q(x_n)\in[0,1]$, we have

$$T(d,q(x_n))=\frac{q(x_n)[q(x_n)-q(x_n^j)]}{\sum_{n=1}^{13}(q(x_n)-q(x_n^j))^2}<\frac{[q(x_n)-q(x_n^j)]^2}{\sum_{n=1}^{13}(q(x_n)-q(x_n^j))^2}<1 \tag{14}$$

That is, when the independent variable $q(x_n)$ changes slightly, the change in the dependent variable d is always smaller than the change in the independent variable $q(x_n)$. It is concluded that the effect of a small change in the independent variable (i.e., the content of each chemical component) on the dependent variable d is small.

So it is said that the sensitivity of our classification results is small and it is a reasonable classification model.

4.4. Correlation and correlation relationship modeling and solving for chemical composition of glass

In this section, the following two problems are solved: analyzing the correlation of chemical components in each glass category and comparing the variability of chemical component correlations between different categories. Our solution is divided into the following three steps.

(1) Keep the relatively important chemical components in the glass and determine whether they have a linear relationship between two of them.

(2) Calculate the correlation coefficients for each of them and check the table to see whether they have significant correlation.

(3) Compare the variability of chemical components between different categories.

Table 2 Correlation analysis of chemical composition of different glass types

Glass Types	Object of comparison	linear relationship(YES/NO)	Correlation coefficient	significant correlation(YES/NO)
Lead barium glass	Two principal components	NO	0.382	YES
	PbO and BaO	NO	0.075	NO
	Al_2O_3 and P_2O_5	NO	0.043	NO
High Potassium Glass	Two principal components	NO	0.001	NO
	K_2O and Al_2O_3	NO	0.476	YES
	Na_2O and K_2O	NO	0.486	YES

Correlation:

Using the results obtained after principal component analysis, as shown in Table 2, the two principal components represent the main components of lead-barium glass and the degree of weathering, respectively, and the correlation analysis of the two aims to determine whether they are correlated, i.e., whether a change in one type of principal component leads to a change in another type of principal component. The scatter plot of the two principal components of lead-barium glass shows that they do not have obvious linear correlation, so the Spearman correlation coefficient, which has stronger generality, is chosen to measure the correlation between them. The correlation coefficient of the two is 0.382, and by consulting the Spearman's correlation coefficient table, we can get that the main components of lead-barium glass and weathering degree have a strong positive correlation. Then we carry out local analysis: according to the analysis of compounds contained in lead-barium glass in the first question, we select several more typical compounds in lead-barium glass to do correlation analysis. It can be obtained that there is no significant correlation between its PbO and BaO, Al_2O_3 and P_2O_5. In the same way, there is no significant correlation between the two main components in high potassium glass, but there are significant correlations between K_2O and Al_2O_3, Na_2O and K_2O. Therefore, in lead-barium glass, the correlation of each variable is reflected in the main components, i.e., the main compound composition of lead-barium glass shows positive correlation with weathering characteristics. In contrast, in high potassium glass, significant correlations exist between K_2O and Al_2O_3, Na_2O and K_2O.

Differences.

(1) For high potassium glass, there is no significant correlation between the two main components, while for lead-barium glass, there is a significant correlation between the two main components.

(2) For high potassium glass, there are mostly significant correlations between the main chemical component K_2O and the rest of each chemical component, so after weathering, the change of K_2O content will lead to the change of the rest of each component, indirectly making the SiO_2 content in a stable state; while the correlations between the main components PbO and BaO of lead-barium glass, and the rest of each component except SiO_2 are weak, making the change of a The changes of one chemical component after weathering can hardly cause the changes of the remaining chemical components, thus leading to the drastic reduction of SiO_2 content after weathering.

5. Conclusion

This paper addresses the practical problem of "compositional analysis and identification of ancient glass products" by organically combining abstract mathematical models, theories and methods with concrete data inspection, analysis and prediction; by filtering, categorizing and analyzing complicated testing data; by innovating and applying classical statistical models. We have realized the verification of the classification and chemical composition correlation of high potassium glass and lead-barium glass, and applied the ideas and methods in the model to the identification of unknown artifact glass types, which has opened up new methods and added new ideas to the investigation of artifacts in archaeological research.

References

[1] Gan Fu Xi, Cheng Huan Sheng, Li Qing Hui. The origin of ancient Chinese glass - a study of the earliest ancient glass in China [J]. Chinese Science (E Series: Technical Science), 2007(03):382-391.

[2] Qiyuan Jiang, Jinxing Xie, Jun Ye, Mathematical Models (3rd ed.) Beijing, Higher Education Press, 2006.9.12

[3] Li QH, Huang KJ, Li F, Gan FX. Chemical composition of a group of ancient Warring States glass samples excavated in China[J]. DOI:10.16334/j.cnki.cn31-1652/k.2006.02.002.