

Sub-classification of glass products based on discriminant function

Wenbo Wang *, Xiaosen Hou, Wuchen Jin

School of Materials Science and Engineering, Yanshan University, Qinhuangdao 066004, China

* Corresponding author: kfwangwb@126.com

Abstract. In this paper, a discriminant function is established for the coarse classification of the known glass, and the sub-classification result of the high-potassium glass is as follows: the $K_2O-Na_2O-SiO_2$ glass, the $K_2O-CaO-SiO_2$ glass, the $K_2O-CuO-SiO_2$ glass, the sub-classification result of lead-barium glass are $PbO-BaO-Na_2O-SiO_2$ glass, the $PbO-BaO-Al_2O_3-SiO_2$ glass, the $PbO-BaO-SrO-SiO_2$ glass, the $PbO-BaO-CaO-SiO_2$ glass. The subtaxonomic results correspond to the color and ornamentation, which indicates the rationality of subtaxonomic results. Sensitivity analysis showed that the sensitivity was small and the model was robust. It provides a reference for the subsequent study of the weathering and composition of ancient silicate glass.

Keywords: Glass chemical composition, Discriminant function, Subclassification.

1. Introduction

China's glass development history has been more than 2,000 years, China's ancient glass production process has reached a certain level, of which silicate glass occupies the main part. This paper hopes to provide some guidance for the digital research and classification of contemporary silicate glass by studying the sub-classification of ancient glass. In the past research, scholars tried to use some methods to analyze the chemical composition of ancient glass, such as LA-ICP-MS (laser ablation-Inductively coupled plasma-mass spectrometry) [1-5] (proton-Particle Induced Gamma-ray Emission) [6-9], XRF (X-ray fluorescence spectrometry) [10, 11], but because of the variety of chemical elements in ancient glass and the complex distribution of each element content, few scholars have systematically established mathematical models and conducted sub-classification research on the chemical composition of ancient glass. The classification of unknown ancient glass samples belongs to discriminant analysis in statistics. Nowadays, discriminant analysis has become an important theoretical basis in data mining, machine learning, pattern recognition and other application fields. In this paper, the new ancient glass samples are classified by the discriminant function and criterion based on the chemical composition analysis.

2. Materials and methods

2.1. Data acquisition and preprocessing

The data source of this paper is Question C (<http://www.mcm.edu.cn/>) of Undergraduate Group of National University Mathematical Contest in Modeling in 2022.

Accumulate the components of each sampling point in the data to obtain the accumulated components and parts as shown in Table 1:

Table 1. Cumulative sum of components at sampling points

Sampling point number	sum of components
01	97.61%
02	99.89%
...	...
57	92.47%
58	98.76%

Among them, the component sum of No.15 and No.17 is 79.47% and 71.89% respectively, which are both less than 85%, so they are eliminated, and the sample points missing in the "color" item in Table 1 are deleted (No.19, 40, 48 and 58 respectively). On this basis, the data corresponding to the unweathered sample points are screened out.

For the high potassium silicate glass, the content of sodium ion is used as the criterion for the first classification. At the same time, because the content of magnesium ion and strontium ion is relatively small compared with calcium ion and copper ion, the representativeness is poor, and phosphorus pentoxide is the impurity component in the cultural relics, which is not the main factor affecting the classification of cultural relics. Therefore, the content of calcium ion and copper ion is used for the second classification of high potassium silicate glass.

The sodium ion and calcium ion are use as that first and second classification criteria, and the strontium and aluminum ions are use for the third sub-classification.

2.2. Method Introduction

The idea of discriminant analysis is to define the center point of each classification, and calculate the distance from the sample point to the center point. If the sample point to be discriminated is close to a certain center, it belongs to the classification.

2.2.1. Determination of distance discriminant function

The classification index of discrimination is recorded as a 2-dimensional vector: $x = (X_1, X_2)^T$, the population center is represented by u , the symbol of covariance matrix is Σ , and considering the correlation between two indexes and the dispersion of values, the distance is Mahalanobis distance:

$$d(x, X) = \sqrt{(x - u)^T \Sigma^{-1} (x - u)} \quad (1)$$

The expectations of X_1 and X_2 are u_1 and u_2 . For the given sample, if $d^2(x, X_1) \geq d^2(x, X_2)$, the sample can be considered to belong to the type represented by population 1. Record that:

$$d^2(x, X_2) - d^2(x, X_1) = (x - u_2)^T \Sigma_2^{-1} (x - u_2) - (x - u_1)^T \Sigma_1^{-1} (x - u_1) \quad (2)$$

According to the above formula can be obtained:

$$F(x) = \frac{1}{2} (d^2(x, X_2) - d^2(x, X_1)) \quad (3)$$

$F(x)$ is the distance discriminant function of the required solution. When the covariance matrix is equal to a fixed value:

$$\begin{cases} F(x) = b^T (x - u) \\ b = \Sigma^{-1} (u_1 - u_2) \\ u = \frac{1}{2} (u_1 + u_2) \end{cases} \quad (4)$$

In this case, $F(x)$ is a linear function.

2.2.2. Estimation and Test of Population Statistics

The expected value vector and covariance matrix of the population need to be estimated. The estimated relationship between the expected value vector and covariance matrix of the population is as follows:

$$\begin{cases} u_1 \approx \bar{x}_1 = \frac{1}{n_1} \sum_{k=1}^{n_1} x_{1k} \\ u_2 \approx \bar{x}_2 = \frac{1}{n_2} \sum_{k=1}^{n_2} x_{2k} \\ \Sigma_1 \approx S_1 = \frac{1}{n_1-1} \sum_{k=1}^{n_1} (x_{1k} - \bar{x}_1)(x_{1k} - \bar{x}_1)^T \\ \Sigma_2 \approx S_2 = \frac{1}{n_2-1} \sum_{k=1}^{n_2} (x_{2k} - \bar{x}_2)(x_{2k} - \bar{x}_2)^T \end{cases} \quad (5)$$

In order to solve the final discriminant function, it is necessary to do that the covariance consistency test is carried out on the two data populations (two glass sub-classification populations) so that analyzing whether the covariance matrix is the same or not. Hypothesis testing is:

$$H_0: \Sigma_1 = \Sigma_2 \quad H_1: \Sigma_1 \neq \Sigma_2 \quad (6)$$

According to statistical knowledge, the corresponding statistic for this hypothesis test should be:

$$L^* = (1 - c)L \sim \chi^2(3) \quad (7)$$

Among them:

$$\begin{cases} L = (n_1 + n_2 - 2) \ln |S_W| - (n_1 - 1) \ln |S_1| - (n_2 - 1) \ln |S_2| \\ c = \frac{13}{18} \left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1} + \frac{1}{n_1 + n_2 - 2} \right) \end{cases} \quad (8)$$

For a given test level α , the probability $p_0 = P(L^* > \chi^2(3))$ is calculated, if $p_0 < \alpha$, the null hypothesis is rejected and the population should be classified using a nonlinear discriminant function. Otherwise, accept the null hypothesis and classify the population using a linear discriminant function.

3. Results

3.1. Establishment of Sub-differentiation Discriminant Model for High Potassium Silicate Glass

3.1.1. Primary Subclassification by Sodium Content

Through screening and preliminary observation data, it is observed that the distribution of sodium oxide content in high-potassium non-weathered glass cultural relics has obvious differentiation.

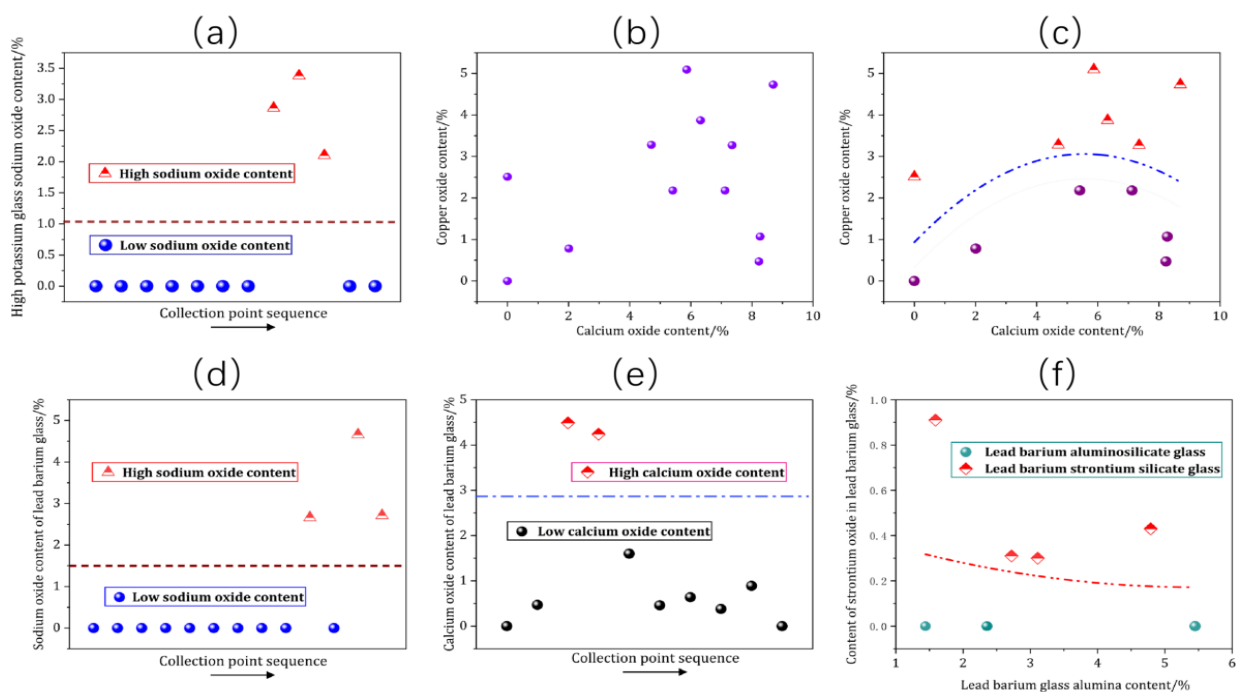


Figure 1. (a) Scatter plot of sample data for high potassium sodium silicate composition. (b) Scatter plot of glass composition sample data. (c) Distance discrimination curve and discrimination result diagram. (d) Scatter plot of sample data for composition of sodium silicate. (e) Scatter plot of sample data of glass lime composition. (f) Sample data scatter plot and discrimination result plot.

According to the literature, the classification methods of ancient glass are mainly as follows:
 1. soda-lime-silicate glass; 2. potassium-lime-silicate glass; 3. Lead-barium-silicate glass

According to the classification method in the references and in combination with Figure.1 (a), it can be seen that: the distribution of sodium oxide content in high potassium non-weathering glass cultural relics has obvious differentiation, according to which the high potassium non-weathering glass cultural relics are classified for the first time.

The results of the primary sub-classification are sodium potash silicate glass and non-sodium potash silicate glass, and the collection points belonging to sodium potash silicate glass are:16, 13, 14.

3.1.2. Secondary subclassification according to divalent metal ion content

The classification index for distinguishing glass is recorded as 2-dimensional vectors $x = (x^{(1)}, x^{(2)})^T$, where $x^{(1)}, x^{(2)}$ are the contents of CaO and CuO respectively, and Figure.1 (b) is the distribution of sample points under this classification index.

Perform the covariance consistency test on the two data populations. Given the test level $\alpha = 0.05$, $p_0 = P(L^* > \chi^2(3))$, the program calculates: $u_1 = (5.1733, 1.1133)$, $u_2 = (5.4917, 3.7917)$, $|S_1| = 12.7180$, $|S_2| = 10.0189$, $c = 0.2167$, $L = 0.0710$, $p_0 = P(L^* > \chi^2(3)) \approx 0.0011 < 0.05$, therefore rejecting the null hypothesis, the non-linear discriminant function should be used to classify the glass for CaO and CuO, solving the covariance matrix, resulting in the non-linear discriminant function:

$$F(x) = -97.840 + 28.7859x_{(1)} - 2.6991x_{(1)}^2 + 13.2391x_{(2)} - 2.6991x_{(2)}^2 \quad (9)$$

If $F(X) < 0$, the glass can be classified as $K_2O-CaO-SiO_2$ glass (potassium-lime silicate glass), otherwise, it is $K_2O-CuO-SiO_2$ glass (potassium-copper silicate glass). The classification results of sample points according to discriminant function are shown in Figure 1 (c).

The collection points belonging to potassium-lime silicate glass are 04 and 06 parts 2, 03 parts 1 and 18. The collection points belonging to potassium-copper silicate glass are:03 site 2, 01, 21, 05, 06 site 1.

3.2. Establishment of subdifferentiation discriminant model for lead-barium silicate glass

3.2.1. Primary and secondary sub-classification according to sodium and calcium content

Similar to the analysis in 3.1.1, the distribution of sodium oxide content of lead-barium non-weathered glass cultural relics obtained in combination with Figure.1 (d) has relatively obvious differentiation, and the primary sub-classification of lead-barium non-weathered glass cultural relics can be carried out accordingly.

The primary sub-classification results are: lead barium sodium silicate glass and lead barium non-sodium silicate glass, among which the lead barium sodium silicate glass is:45, 55, 47.

In the same way, the distribution of CaO content in lead-barium non-sodium glass obtained from Figure 1 (e) also has obvious differentiation, which can be used to classify the lead-barium non-weathered glass relics for the second time.

3.2.2. On the basis of strontium, aluminum oxide content category three times

In consideration of strontium and aluminum oxide content for sub-classification type discrimination, that discrimination model in the high-potassium silicate glass is use for solving.

Given the inspection level $\alpha = 0.05$, $p_0 = P(L^* > \chi^2(3))$, the program calculates: $u_2 = (3.0525, 0.4875)$, $u_1 = (2.9, 0)$, $|S_1| = 3.0761$, $|S_2| = 2.1879$, $c = 0.3611$, $L = 0.0866$, $p_0 = P(L^* > \chi^2(3)) \approx 0.00012 < 0.05$, therefore, the null hypothesis is rejected, and the covariance matrix also fails the consistency test. Use the quadratic discriminant function to obtain the final discriminant function by using the sample data and the corresponding formula of (4):

$$F(x) = -10.6829 + 5.40665x_{(1)} - 0.8822x_{(1)}^2 + 2.3603x_{(2)} - 0.8822x_{(2)}^2 \quad (10)$$

If $F(x) > 0$, the glass can be classified as $PbO - BaO - Al_2O_3 - SiO_2$ glass (lead barium aluminosilicate glass), otherwise $PbO - BaO - SrO - SiO_2$ glass (lead barium strontium silicate glass). The classification results of the sample points according to the discriminant function are shown in Figure 1 (f).

3.3. Model Checking and Result Reasonableness Interpretation

3.3.1. Justification of classification results

Table 2. The embodiment of different types of glass color, decorative pattern

Type	Subclassification type	Color	Ornamentation	Classification Results
high-potassium glass	the K ₂ O-Na ₂ O-SiO ₂ glass	Mainly light blue	C	16, 13, 14
	the K ₂ O-CaO-SiO ₂ glass	Mainly turquoise	Mainly C	04, 06 parts 2, 3 part 1, 18
	the K ₂ O-CuO-SiO ₂ glass	Mainly turquoise	Mainly A	03 part 2, 01, 21, 05, 06 part 1
lead-barium glass	PbO-BaO-Na ₂ O-SiO ₂ glass	/	/	45, 55, 47
	the PbO-BaO-Al ₂ O ₃ -SiO ₂ glass	Mainly dark blue	Mainly A	30 parts 1 and 2
	the PbO-BaO-SrO-SiO ₂ glass	/	/	33, 32, 35, 20
	the PbO-BaO-CaO-SiO ₂ glass	Mainly light blue	Mainly A	24, 46, 37, 31

Under the sub-classification, the analysis of Table 2 shows that except for lead-barium-sodium silicate glass and strontium-non-calcium-non-lead-barium-sodium silicate glass, other sub-classification types have corresponding colors and ornamentation, and have obvious differences for different types. By comparison with the attached data, the following conclusions can be drawn: The potassium-copper silicate glass tends to A ornamentation and blue-green in color, while the potassium-calcium silicate glass tends to C ornamentation, which indicates that the classification of high-potassium non-weathering glass relics based on monovalent sodium ion and divalent calcium ion is reasonable.

The classification indirectly combines the composition with the external characterization, which has a certain guiding significance for analyzing the composition of glass cultural relics through the external characteristics.

3.3.2. model checking

In order to show that the discriminant classification effect is better, sensitivity analysis is carried out, and the discriminant model of high potassium silicate glass sub-differentiation is taken as an example.

When the discriminant function is equal to zero, the corresponding implicit function image is the curve dividing two parts. If the value of the discriminant function is changed, the sample points can still be divided correctly, which indicates that the discriminant model has poor sensitivity and good stability. When changed, it causes the boundary to shift, thus drawing a two-dimensional curve. If the band area is large without affecting the correct discrimination, it indicates that the sensitivity is poor and the discriminant classification model is effective. To solve for the maximum two-dimensional interface, the discriminant function is changed as follows:

$$F(x) = -97.8940 + 28.7859x_{(1)} - 2.6991x_{(1)}^2 + 13.2391x_{(2)} - 2.6991x_{(2)}^2 + K_0 \quad (11)$$

When K_0 changes, the dividing line changes accordingly, and the area S_0 swept by the dividing line is the research object. According to the data analysis, the four coordinates most sensitive to the

dividing line are:(5.87, 5.09),(7.12, 2.18),(7.35, 3.27),(4.71, 3.28), and substituting them into equation (2), the following results are obtained:

$$\begin{cases} K_{01} = 24.4649 \\ K_{02} = 13.7336 \\ K_{02} = 17.6991 \\ K_{03} = 7.8033 \end{cases} \quad (12)$$

Thus, the classification discrimination line can be changed into classification discrimination curve band, and the discrimination curve band diagram Figure.2

$$\begin{cases} F(x) = -97.8940 + 28.7859x_{(1)} - 2.6991x_{(1)}^2 + 13.2391x_{(2)} - 2.6991x_{(2)}^2 + K_0 \\ K_0 \in [7.8033, 24.4649] \end{cases} \quad (13)$$

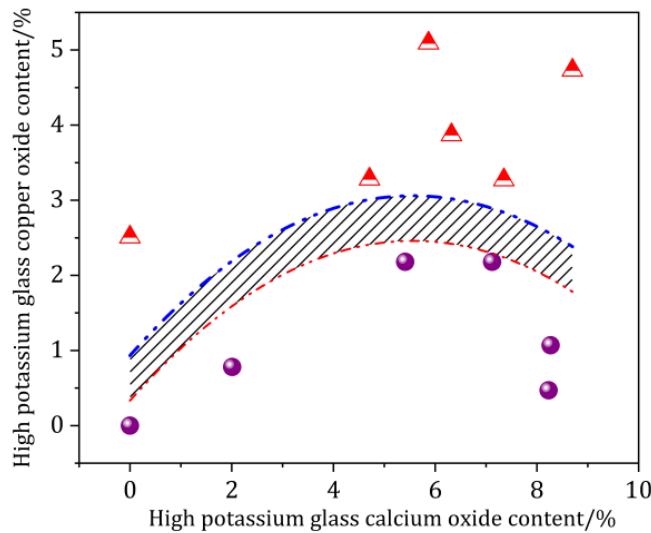


Figure 2. Discriminant band diagram of high-potassium glass

It can be seen from Figure 2 that the classification and discrimination curve has a large area, and the width of the band has even reached one sixth of the length of the variable region. The sensitivity is small, that is, the model is relatively robust. For other sub-classification models, this judgment method is also applicable.

4. Conclusion

Based on the rough classification, the discriminant function is established, and the sub-classification results of high-potassium glass are as follows: $K_2O - Na_2O - SiO_2$ glass, $K_2O - CaO - SiO_2$ glass and $K_2O - CuO - SiO_2$ glass, lead barium glass sub-classification results are: $PbO - BaO - Na_2O - SiO_2$ glass, $PbO - BaO - Al_2O_3 - SiO_2$ glass, $PbO - BaO - SrO - SiO_2$ glass and $PbO - BaO - CaO - SiO_2$ glass. The rationality of subclassification was verified by comparing the results of subclassification with color and ornamentation. Sensitivity analysis showed that the sensitivity was small and the model was robust. The construction of this model provides some technical support and theoretical support for the classification of ancient glass chemical composition.

References

- [1] Cagno S, Favaretto L, Mendera M, Izmer A, Vanhaecke F, Janssens K (2012) Evidence of early medieval soda ash glass in the archaeo-logical site of San Genesio (Tuscany). *J Archaeol Sci* 39 (5): 1540 – 1552.
- [2] Dussubieux L, Golitko M, Ryan Williams P, Speakman RJ (2007) In Glascock MD, Speakman RJ, Popelka-Filcoff RS (Eds.) *Archaeological chemistry, analytical techniques and archaeological interpretation* 968: 349 – 363.
- [3] De Raedt I, Janssens K, Veeckman J, Vincze L, Vekemans B, Jeffries TE (2001) Trace analysis for distinguishing between Venetian and façon-de-Venise glass vessels of the 16th and 17th century. *J Anal Atom Spectrom* 16 (9): 1012 – 1017.
- [4] Panighello S, Orsega EF, van Elteren JT, Šelih VS (2012) Analysis of polychrome Iron Age glass vessels from Mediterranean I, II and III groups by LA-ICP-MS. *J Archaeol Sci* 39 (9): 2945 – 2955.
- [5] Wagner B, Nowak A, Bulska E, Kunicki-Goldfinger J, Schalm O, Janssens K (2008) Complementary analysis of historical glass by scanning electron microscopy with energy dispersive X-ray spectroscopy and laser ablation inductively coupled plasma mass spectrometry. *Microchim Acta* 162: 415 – 424.
- [6] Gomez-Tubio B, Ontalba-Salamanca MA, Ortega-Feliu I, Res-paldiza MA (2006) PIXE-PIGE analysis of late roman glass fragments. *Nucl Instrum Methods B* 249: 616 – 621.
- [7] Smit Z, Milavec T, Fajfar H, Rehren Th, Lankton JW, Gratuze B (2013) Analysis of glass from the post-Roman settlement Tonovcov grad (Slovenia) by PIXE-PIGE and LA-ICP-MS. *Nucl Instrum Methods Phys Res B* 311: 53 – 59.
- [8] Dasari KB, Chhillar S, Acharya R, Ray DK, Behera A, Lakhmana Das N, Pujari PK (2014) Simultaneous determination of Si, Al and Na concentrations by particle induced gamma-ray emission and applications to reference materials and ceramic archaeological artifacts. *Nucl Instrum Methods Phys Res B* 339: 37 – 41.
- [9] Sastri C S, Sauvage T, Blondiaux G, et al. Analysis of Na, Mg, Al and Cl in archaeological glass and pottery: comparison of PIGE with low energy deuteron activation analysis [J]. *Journal of Radioanalytical and Nuclear Chemistry*, 2020, 324 (1): 159 - 167.
- [10] Sastri C S, Sauvage T, Blondiaux G, et al. Analysis of Na, Mg, Al and Cl in archaeological glass and pottery: comparison of PIGE with low energy deuteron activation analysis [J]. *Journal of Radioanalytical and Nuclear Chemistry*, 2020, 324 (1): 159 - 167.
- [11] Yatsuk O, Ferretti M, Gorghinian A, et al. Data from Multiple Portable XRF Units and Their Significance for Ancient Glass Studies [J]. *Molecules*, 2022, 27 (18): 6068.