

Component analysis and identification of glass artifacts based on logistic regression and K-means clustering

Jiangbiao Xiao ^{1,*}, Junjie Kong ², Siwei Leng ³

¹ College of Water Conservancy and Ecological Engineering, The Nanchang Institute of Technology, Nanchang, China

² School of International Education, The Nanchang Institute of Technology, Nanchang, China

³ School of Information Engineering, The Nanchang Institute of Technology, Nanchang, China

* Corresponding Author Email: 2839095328@qq.com

Abstract. The Silk Road was a channel of cultural exchange between China and the West in ancient times, and ancient glass relics are valuable physical evidence witnessing the history of early trade between China and the West. Early glass was often made into bead-shaped ornaments in the West Asian and Egyptian regions and imported into China, and ancient glass in China was made locally after absorbing its technology. Although the two types of glass products are similar in appearance, their chemical compositions are not identical. This paper focuses on the influence of weathering on high potassium glass and lead-barium glass and the analysis and identification of the composition of the two types of glass. Based on the logistic model to derive their correlation factors, we determine the correlation between weathering and glass type, decoration and color on the surface of cultural relics. Using stepwise regression analysis, the basis for the classification of high potassium glass and lead-barium glass is derived.

Keywords: Logistic model; stepwise regression; correspondence analysis; glass weathering.

1. Introduction

The Silk Road was one of the most important mediums of exchange between Eastern and Western cultures in ancient times, and glass can be used as an important credential of exchange. Although glass looks basically the same from the outside, the chemical composition of Chinese glass and Western glass is not the same [1]. Because ancient glass is susceptible to weathering due to bad weathering factors, during this weathering time period, a large number of chemical changes are likely to occur, resulting in changes in the distribution of elements, which has a huge impact on our discernment [2-4].

The main raw material of glass is quartz sand, the core chemical composition of which is silicon dioxide (SiO₂). In the process of glass production we need to add a flux because we want to lower the melting point of pure quartz. The flux is usually grass ash, natural soda ash, etc. The stabilizer is limestone, which is converted to calcium oxide after calcination. After analyzing two main types of high potassium glass and lead-barium glass [5].

This paper mainly explores the relationship between the surface weathering of glass artifacts and their glass type, decoration and color. Based on the type of glass, explore the pattern of whether the artifacts have weathered chemical composition data on their exterior. For high potassium glass and lead-barium glass, the appropriate chemical composition is selected as a classification criterion for more refined classification, providing the method, basis, and results of the classification.

2. Model assumptions and notation

2.1. Assumptions [6]

1. Assuming that there are no other chemical reactions in the artifact samples.
2. Assuming that no other contamination of the artifact sample affects the color.
3. Assume that only the 14 oxides listed are present in the artifact sample.

2.2. Notations

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

Table 1. Notations

Symbols	Description
α	Chemical content before weathering
λ	Chemical content after weathering
x_1	Decoration A
x_2	Decoration B
x_3	Texture C
x_4	Type high potassium
x_5	Type lead barium
x_6	Color blue-green
x_7	Color violet
x_8	Color light blue
x_9	Color dark green
x_{10}	Color light green
x_{11}	Color black
x_{12}	Color green
x_{13}	Color is empty
x_{14}	Surface weathering
p	Corresponding matrix
F	Column contour coordinates

3. Model construction and solving

3.1. Logistic model building

Tep 1: Assignment solution

First, the 14 parameters (A, B, C, high potassium, lead barium, blue-green, purple, light blue, dark green, light green, black, green, empty, surface weathering) in the ornamentation, category, color and surface weathering were set in order, and then the data were processed for the representative data, resulting in the first 13 parameters, each artifact has an ornamentation, a category, a color, these three characteristics corresponding to the three parameters [7]. The remaining 10 parameters were identified as 1, and the 14th parameter was identified as 0. The 14th parameter with weathering was 1, and the opposite was 0 (e.g., artifact number 01, ornament C, type high potassium, color blue-green, and no weathering on the surface, was output as 00110100000000 according to our regulations). Then for 58 representative samples of artifacts according to the rules to make Table 2.

Table 2. Assigning values to 58 samples of cultural relics

Artifact Number	A	B	C	High Potassium	Lead Barium	Blue-Green	Purple	Pale Blue	Dark Green	light green	Black	Green	Empty
01	0	1	0	1	0	0	0	0	0	0	0	0	0
02	0	1	1	0	1	0	0	0	0	0	0	0	0
03	0	0	0	1	0	0	1	0	0	0	0	0	1
04	0	0	1	0	1	0	0	0	0	0	0	0	0
05	0	0	1	0	1	0	0	0	0	0	0	0	0
06	0	0	1	0	1	0	0	0	0	0	0	0	0
07	0	0	1	0	1	0	0	0	0	0	0	0	0
08	1	0	1	0	1	0	0	0	0	0	0	0	1

Step 2: Model building

The basic form of the logistic model is [8].

$$P(Y=1|x_1, x_2, \dots, x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}. \quad (1)$$

However, in the practical application of the model, often instead of regressing P directly, a monotonic continuous probability function is first defined such that

$$\pi = P(Y=1|x_1, x_2, \dots, x_k), \quad 0 < \pi < 1. \quad (2)$$

Thus the logistic model can also be deformed as:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k, \quad 0 < \pi < 1. \quad (3)$$

The data of 13 parameters and surface weathering related factors obtained based on the logistic regression model are shown in Table 3.

Table 3. Table of correlation factors between type, ornamentation and color and surface weathering

The first 13 parameters	Related factors factors	Correlation with surface weathering
A	0.9307	Positive correlation
B	-2.2790	Negative correlation
C	0	No correlation
High Potassium	-1.6912	Negative correlation
Lead Barium	0	No correlation
Blue-green	1.8547	Positive correlation
Violet	0.1950	Positive correlation
light blue	-1.0672	Negative correlation
dark green	-0.1826	Negative correlation
light green	-0.6453	Negative correlation
Black	-1.4334	Negative correlation
green	0.6192	Positive correlation
empty	-2.1698	Negative correlation

According to the data of correlation factors in the table, the correlation between surface weathering of artifacts and their glass type, decoration and color can be determined. Positive correlations with surface weathering are A, blue-green, purple, and green; negative correlations are B, high potassium, light blue, dark green, light green, black, and empty; and C and lead-barium are not correlated with surface weathering.

Step 3: chi-square test

The chi-square is calculated by.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}, \quad (4)$$

The numerator of the formula represents the deviation of the actual value from the expected value, while the denominator is the process of standardization. Therefore, the smaller the chi-square, the closer the observed value is to the expected value theoretical value, indicating that the two variables are more consistent with the chi-square distribution, which presupposes that the variables are relatively independent, so the smaller the value of the chi-square, the more independent the two variables are from each other.

The basic process of the chi-square test is as follows.

(1) establish a hypothesis test, the original hypothesis is generally independent of each other between the variables.

(2) Calculate the expected frequency.

(3) Calculate the chi-square value by bringing in the chi-square statistical formula.

(4) Calculate the degrees of freedom.

(5) Check the table and compare the chi-squared values or p-values.

3.2. Effect of weathering on high potassium glass and lead-barium glass

Step 1: Data processing

According to the requirements, the data of the artifact samples that are cumulative and do not meet the conditions for the composition ratio are invalid data, the conditions are shown in Table 4, the data shown as NaN in Matlab are replaced with 0, and the data are filtered using the cumulative method, and the invalid data are successfully excluded. Next, according to the requirements of the topic according to the type of analysis, this question has a total of two categories of glass were high potassium glass and lead barium glass, so the two types of glass weathered and unweathered in two cases were compared in different types of chemical materials in the chemical composition of the proportion, using the comparison method to explore whether the surface of the heritage samples weathered chemical composition of the number of laws [9,10].

Because the superiority of the line graph is that it can not only show the number, but also can reflect the change of things in different time situations, so the line graph was finally used to compare the effect of weathering on the chemical composition within the samples. See Figure 1.

Table 4. Judgment conditions for valid data

Component ratio accumulation and s	Is the data valid
$s \in [85\%, 105\%]$	Yes
$s \notin [85\%, 105\%]$	No

Step 2: Determine the category interval according to the icon

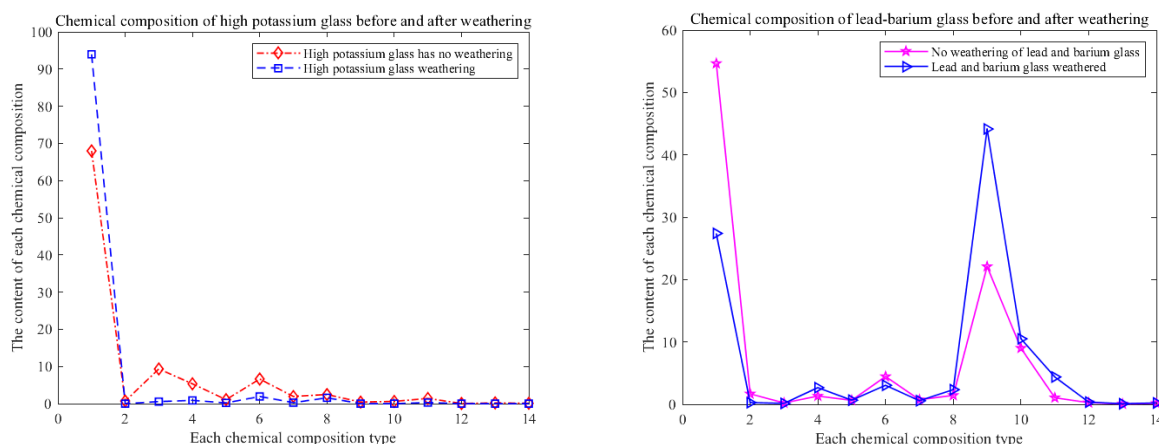


Figure 1. Chemical composition of high potassium glass and lead-barium glass before and after weathering

As shown in Figure 1, it can be seen that high potassium glass without weathering has only 70% chemical composition when there is only one chemical composition type, while the weathered one is close to 100%. In the case of only three types of chemical composition, the chemical content is 10%, while the weathering is close to 0%. In the case of only four types of chemical composition, the chemical content is 5%, while the weathering is close to 0%. In the case of 6 types of chemical composition, the chemical content is 10%, while the weathering is close to 5%. It can be seen that barium lead glass without weathering has 55% chemical composition when there is only one chemical composition type, while the weathered one is close to 30%. When there are only 9 types of chemical composition, the chemical content is 25%, while the weathering is nearly 50%. The chemical content was 3% when there were only 11 types of chemical components, while the weathered ones were close to 10%.

When comparing the unweathered high potassium glass with the weathered high potassium glass, the content of four chemical components, SiO₂, K₂O, CaO, and Al₂O₃, differed greatly. The statistical law of the weathered or unweathered chemical content on the surface of the artifact samples was obtained: to determine whether the surface of the high-potassium glass is weathered or not, the content of the four chemical components, SiO₂, K₂O, CaO, and Al₂O₃, is mainly judged, and if the content of SiO₂ is higher than the normal value, the surface of the high-potassium glass is weathered, and vice versa, it is unweathered. If K₂O, CaO, and Al₂O₃ are lower than normal, the surface of the high potassium glass is weathered, and if not, the surface is unweathered.

If the content of SiO₂, PbO, P₂O₅ is lower than the normal value, the surface of the lead-barium glass has been weathered, and vice versa, the surface of the lead-barium glass has not been weathered.

3.3. The basis of glass division based on stepwise regression analysis

Step 1: Model solving

For the problem, Bidirectional elimination: stepwise screening method is used.

When a variable is introduced, we first check whether the variable makes a significant change in the model (F-test), and if a significant change occurs, we then test all the variables, and when the original introduced variable no longer changes significantly due to the introduction of the later added variables, we eliminate this variable to ensure that only significant variables are included in the regression equation before each new variable is introduced. The regression equation contains only significant variables before each new variable is introduced, until neither significant explanatory variables are selected into the regression equation, nor insignificant explanatory variables are removed from the regression equation, and finally an optimal set of variables is obtained. To establish the one-variable regression equation for each independent variable and the dependent variable.

(1) Establishing a one-variable regression equation for each independent variable and the dependent variable $y = a_i X_i + b_i, i = 1, 2, \dots, m$

(2) Calculate the test statistic F for the regression coefficients in each of the m one-dimensional regression equations, and find the maximum value of $F_{k_1}^1 = \max\{F_1^1, F_2^1, \dots, F_m^1\}$

If $F_{k_1}^1 \leq F_\alpha(1, n-2)$, stop Screening

(3) Establish a binary regression equation for the independent variable group $(x_1, x_2), (x_1, x_3), \dots, (x_1, x_m)$ and the dependent variable respectively, (at this time x_1 is x_{k_1} in step 2) calculate the regression coefficient test statistic F for $x_1, x_2, x_3, \dots, x_m$ in the equation, and take

Take $F_{k_2}^2 = \max\{F_2^2, F_3^2, \dots, F_m^2\}$, if $F_{k_2}^2 \leq F_\alpha(1, n-2-1)$ then stop filtering, otherwise x_{k_2} will be selected into the variable set, at this time x_{k_2} will be seen as x_{k_1} . So iterate until the maximum F value of the independent variable is less than the critical value, then the regression equation is the optimal regression equation.

Step 2: Data processing

The data were processed again, and the category to which the artifact numbers belonged was added to the column in Form 2 in order according to the rule that high potassium glass is 1 and lead-barium glass is 0. Then, using the stepwise regression analysis method, the final run results were obtained after running and clicking Next Step, and the p -values could be obtained below 0.05, which can indicate that SiO_2 , Na_2O , MgO , CuO , and PbO are the basis for the classification of high potassium glass and lead-barium glass. PbO is the basis for the division of high potassium glass and lead-barium glass, so the investigation concluded that the division pattern of high potassium glass and lead-barium glass is related to these five oxides.

4. Conclusion

This paper focuses on the effects of high potassium glass and lead-barium glass and the compositional analysis and identification of the two types of glass. Based on the logistic model to derive correlation factors, the correlation between surface weathering and glass type, decoration and color of cultural relics was determined. Using stepwise regression analysis, the basis for the classification of high-potassium glass and lead-barium glass was derived.

In this paper, according to the logistic regression model, the proportion of each component can be predicted accurately, which is easy to operate and has high available value.

When dealing with high potassium glass, the data provided by high potassium glass is too little, which leads to the clustering is not too good and does not form relatively concentrated clusters.

The discriminant analysis model can also be used to come to a determination of the different species species for a detailed analysis of the various substances to be classified.

References

- [1] Li M. Preparation of Fionse products and study on lead-barium glass in Warring States, Qin and Han Dynasties [D]. Beijing University of Chemical Technology, 2014. (in Chinese) https://blog.csdn.net/qq_40298902/article/details/101225330
- [2] Zhou Xuwen, Gao Xuxu, Yu Changwen, Han Shiru, Xu Qihui. Quantitative prediction of air pollution potential based on Bayes discriminant analysis [J]. Meteorological Monthly, 202, 48(08):1032-1042.
- [3] Chen, R.E.. Analysis of the rationality of statistical grouping [J]. Statistics Education, 2000(05):9-11.
- [4] Miriam Bruhn. A Tale of Two Species: Revisiting the Effect of Registration Reform on Informal Business Owners in Mexico[M]. :2012-02-01.
- [5] Shoukui Si, Xiqing Sun. Mathematical modeling algorithms and applications [M]. Beijing: University of Defense Technology Press, 2016-2.

- [6] Zhong L. Financial risk early warning analysis of Small and medium-sized enterprises based on factor analysis and Logistic regression method [J]. Journal of Commercial Accounting,2021(21):66-69.
- [7] Wang C Y, Tao Y. Weathering of silicate glass [J]. Journal of the Chinese Ceramic Society, 2003(01):78-85. (in Chinese)
- [8] Tao Ying, Bo Xuwei, Wang Chengyu. Barium phosphate optical glass weathering research [J]. Journal of silicate bulletin, 1996 (04) : 53-57, DOI: 10.16552 / j.carol carroll nki issn1001-1625.1996.04.013.
- [9] Song Bin. On the Correlation between statistical correlation and regression analysis [J]. Modern Vocational Education,2022(21):97-99.
- [10] Pfau Thilo, Bolt David M., FiskeJackson Andrew, Gerdes Carolin, Hoenecke Karl, Lynch Lucy,Perrier Melanie, Smith Roger K. W.. Linear Discriminant Analysis for Investigating Differences in Upper Body Movement Symmetry in Horses before/after Diagnostic Analgesia in Relation to Expert Judgement[J]. Animals,2022,12(6).